

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

BEST AVAILABLE COPY

INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

-1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.

-1- SUBCELLULAR LOCATION: Secreted.

-1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announcement/](http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch) or send an email to license@isb-sib.ch).

RESULT 15
 INS2_MOUSE
 ID INS2_MOUSE STANDARD; PRT; 110 AA.
 AC P01326;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Insulin 2 precursor.
 GN INS2 OR INS-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87169768; PubMed=3104603;
 RA Wentworth B.M., Schaefer I.M., Villa-Komaroff L., Chirgwin J.M.;
 RT "Characterization of the two nonallelic genes encoding mouse
 preproinsulin.";
 RL J. Mol. Evol. 23:305-312(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NON;
 RX MEDLINE=90372989; PubMed=2397023;
 RA Sawa T., Ohgaku S., Morioka H., Yano S.;
 RT "Molecular cloning and DNA sequence analysis of preproinsulin genes
 in the NON mouse, an animal model of human non-obese, non-insulin-
 dependent diabetes mellitus.";
 RL J. Mol. Endocrinol. 5:61-67(1990).
 RN [3]
 RP SEQUENCE OF 25-54 AND 90-110.
 RX MEDLINE=72189455; PubMed=5063718;
 RA Buezli H.F., Glatthar B., Kunz P., Muelhaupt E., Humbel R.E.;
 RT "Amino acid sequence of the two insulins from mouse (Mus musculus).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 353:451-458(1972).
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 DISULFIDE BONDS.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X04724; CAA28433.1; -
 DR PIR; B01592; INMS2.
 DR PIR; A26342; A26342.
 DR PIR; B48172; B48172.
 DR HSP; P01317; IAPH.
 DR MGD; MGI:96573; Ins2.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00276; INSULINA.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism; Signal; Multigene family.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN 2 B CHAIN.
 FT PROPEP 57 87 INSULIN 2 C PEPTIDE.
 FT CHAIN 90 110 INSULIN 2 A CHAIN.
 FT DISULFID 31 96 INTERCHAIN.
 FT DISULFID 43 109 INTERCHAIN.

FT DISULFID 95 100
 SQ SEQUENCE 110 AA; 12364 MW; 3554C8803D24FDAD CRC64;

Query Match 47.6%; Score 50; DB 1; Length 110;
 Best Local Similarity 50.0%; Pred. No. 0.17;
 Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
 Db 91 IVDQCCTSGSLYQLENYCN 110

RESULT 16

INS2_RAT
 ID INS2_RAT STANDARD; PRT; 110 AA.
 AC P01323;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Insulin 2 precursor.
 GN INS2 OR INS-2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Spraue-Dawley; TISSUE=Liver;
 RX MEDLINE=80045035; PubMed=498284;
 RA Lomedico P., Rosenthal N., Efstratiadis A., Gilbert W., Kolodner R.,
 RA Tizard R.;
 RT "The structure and evolution of the two nonallelic rat preproinsulin
 genes.";
 RL Cell 18:545-558(1979).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86310882; PubMed=2427930;
 RA Soares M.B., Schin E., Henderson A., Karathanasis S.K., Cate R.,
 RA Zeitlin S., Chirgwin J., Efstratiadis A.;
 RT "RNA-mediated gene duplication: the rat preproinsulin I gene is a
 functional retroposon.";
 RL Mol. Cell. Biol. 5:2090-2103(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=80240379; PubMed=6249167;
 RA Lomedico P.T., Rosenthal N., Kolodner R., Efstratiadis A.,
 RA Gilbert W.;
 RT "The structure of rat preproinsulin genes.";
 RL Ann. N.Y. Acad. Sci. 343:425-432(1980).
 RN [4]
 RP SEQUENCE OF 25-54 AND 90-110.
 RX MEDLINE=70067613; PubMed=4311938;
 RA Steiner D.F., Clark J.L., Nolan C., Rubenstein A.H., Margoliash E.,
 RA Aten B., Oyer P.E.;
 RT "Proinsulin and the biosynthesis of insulin.";
 RL Recent Prog. Horm. Res. 25:207-282(1969).
 RN [5]
 RP SEQUENCE OF 57-87.
 RX MEDLINE=73061498; PubMed=4640931;
 RA Tager H.S., Steiner D.F.;
 RT "Primary structures of the proinsulin connecting peptides of the rat
 and the horse.";
 RL J. Biol. Chem. 247:7936-7940(1972).
 RN [6]
 RP SEQUENCE OF 57-87, AND REVISIONS.
 RX MEDLINE=72177385; PubMed=4554104;
 RA Markussen J., Sundby F.;
 RT "Rat-proinsulin C-peptides. Amino-acid sequences.";
 RL Eur. J. Biochem. 25:153-162(1972).
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:47 : Search time 57.6471 Seconds
(without alignments)
6.107 Million cell updates/sec

Title: US-09-574-443-7_COPY_1_21
Perfect score: 105
Sequence: 1 NLVQASTQASLYQIYNFDN 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	50.5	21	9	US-09-947-563-1
2	53	50.5	21	10	US-09-853-844-1
3	53	50.5	21	10	US-09-815-229-1
4	53	50.5	30	10	US-09-815-229-16
5	53	50.5	50	9	US-10-066-009A-3
6	53	50.5	50	10	US-09-853-844-4
7	53	50.5	51	9	US-10-028-410-3
8	53	50.5	52	9	US-10-054-873-5
9	53	50.5	54	10	US-09-815-229-13
10	53	50.5	86	9	US-09-878-380-1
11	53	50.5	86	9	US-10-028-410-2
12	53	50.5	86	9	US-10-054-873-4
13	53	50.5	96	9	US-09-947-563-4
14	53	50.5	107	9	US-10-054-873-6
15	53	50.5	110	9	US-09-804-409A-9
16	53	50.5	110	10	US-09-205-658-125
17	53	50.5	110	10	US-09-815-229-3
18	53	50.5	117	10	US-09-280-030-63
19	53	50.5	124	10	US-09-736-611-12

20	53	50.5	124	10	US-09-740-359-12	Sequence 12, Appl
21	53	50.5	124	10	US-09-894-711-12	Sequence 12, Appl
22	53	50.5	124	10	US-09-894-711-18	Sequence 18, Appl
23	53	50.5	125	10	US-09-736-611-10	Sequence 10, Appl
24	53	50.5	125	10	US-09-740-359-10	Sequence 10, Appl
25	53	50.5	125	10	US-09-894-711-10	Sequence 10, Appl
26	53	50.5	130	10	US-09-280-030-62	Sequence 62, Appl
27	53	50.5	144	10	US-09-736-611-6	Sequence 6, Appl
28	53	50.5	144	10	US-09-740-359-5	Sequence 5, Appl
29	53	50.5	146	10	US-09-894-711-5	Sequence 5, Appl
30	53	50.5	147	10	US-09-736-611-8	Sequence 8, Appl
31	53	50.5	147	10	US-09-740-359-7	Sequence 7, Appl
32	53	50.5	150	9	US-10-054-873-7	Sequence 7, Appl
33	50	47.6	21	9	US-09-947-563-7	Sequence 7, Appl
34	50	47.6	46	10	US-09-205-658-132	Sequence 132, App
35	50	47.6	46	10	US-09-205-658-133	Sequence 133, App
36	50	47.6	96	9	US-09-947-563-5	Sequence 5, Appl
37	47	44.8	144	10	US-09-894-711-7	Sequence 7, Appl
38	43	41.0	51	10	US-09-853-844-3	Sequence 3, Appl
39	43	41.0	382	10	US-09-825-414-64	Sequence 64, Appl
40	42	40.0	46	10	US-09-205-658-136	Sequence 136, App
41	42	40.0	604	10	US-09-907-509-2	Sequence 2, Appl
42	41	39.0	46	10	US-09-205-658-135	Sequence 135, App
43	41	39.0	591	10	US-09-815-242-13625	Sequence 13625, A
44	41	39.0	1036	9	US-09-918-508-4	Sequence 4, Appl
45	41	39.0	1092	9	US-10-135-322-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-947-563-1
: Sequence 1, Application US/09947563
: Patent No. US20020156234A1
: GENERAL INFORMATION:
: APPLICANT: Rubroder, Franz-Josef
: Keller, Reinhold
: TITLE OF INVENTION: Improved process for obtaining
: insulin precursors having correctly bonded cystine
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA: US/09/947,563
: APPLICATION NUMBER: 09/134,836
: FILING DATE: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: Leslie McDowell
: REGISTRATION NUMBER: 34,872
: REFERENCE/DOCKET NUMBER: 02481.1600-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 21 amino acids
: TYPE: amino acid
: STRANDEDNESS: single

;
; TOPOLOGY: Linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..21
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-947-563-1

Query Match 50.5%; Score 53; DB 9; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0077;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:111 11 1111: 1:
Db 2 IVEQCCTSCISLYQLENYCN 21

RESULT 2
US-09-853-844-1
; Sequence 1, Application US/09853844
; Patent No. US20020013269A1
; GENERAL INFORMATION:
; APPLICANT: Balschmidt, Per
; APPLICANT: Brange, Veilgaard Jens Jorgen
; TITLE OF INVENTION: Human Insulin Analogues
; FILE REFERENCE: 3343.270-US
; CURRENT APPLICATION NUMBER: US/09/853.844
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 08/965,221
; PRIOR FILING DATE: 1997-11-06
; PRIOR APPLICATION NUMBER: 08/531,842
; PRIOR FILING DATE: 1995-09-21
; PRIOR APPLICATION NUMBER: 08/275,196
; PRIOR FILING DATE: 1994-07-14
; PRIOR APPLICATION NUMBER: 07/976,805
; PRIOR FILING DATE: 1992-11-16
; PRIOR APPLICATION NUMBER: 07/453,445
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: 07/416,218
; PRIOR FILING DATE: 1989-10-02
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-853-844-1

Query Match 50.5%; Score 53; DB 10; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0077;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:111 11 1111: 1:
Db 2 IVEQCCTSCISLYQLENYCN 21

RESULT 3
US-09-815-229-1
; Sequence 1, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDERS
; FILE REFERENCE: PI786RIUS
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17

;
; SEQ ID NO 1
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-229-1

Query Match 50.5%; Score 53; DB 10; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0077;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:111 11 1111: 1:
Db 2 IVEQCCTSCISLYQLENYCN 21

RESULT 4
US-09-815-229-16
; Sequence 16, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDERS
; FILE REFERENCE: PI786RIUS
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 16
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Portion of insulin variant from EP 171,887.
US-09-815-229-16

Query Match 50.5%; Score 53; DB 10; Length 30;
Best Local Similarity 55.0%; Pred. No. 0.011;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:111 11 1111: 1:
Db 11 IVEQCCTSCISLYQLENYCN 30

RESULT 5
US-10-066-009A-3
; Sequence 3, Application US/10066009A
; Patent No. US20020165155A1
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Michelle
; APPLICANT: Ultsch, Mark
; APPLICANT: Vajdos, Felix
; TITLE OF INVENTION: CRYSTALLIZATION OF IGF-1
; FILE REFERENCE: PI869R1
; CURRENT APPLICATION NUMBER: US/10/066,009A
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: US 60/287,072
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/267,977
; PRIOR FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 5
; SEQ ID NO 3
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-066-009A-3

Query Match 50.5%; Score 53; DB 9; Length 50;
Best Local Similarity 55.0%; Pred. No. 0.02;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 2 LVEQASTSQASLYQIYNFDN 21
:|||||||:|:
Db 31 IVEQCCTSIQSLYQLENYCN 50

RESULT 6
US-09-853-844-4
; Sequence 4, Application US/09853844
; Patent No. US20020013269A1
; GENERAL INFORMATION:
; APPLICANT: Balschmidt, Per
; APPLICANT: Brange, Veilgaard Jens Jorgen
; TITLE OF INVENTION: Human Insulin Analogues
; FILE REFERENCE: 3343.270-US
; CURRENT APPLICATION NUMBER: US/09/853,844
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 08/965,221
; PRIOR FILING DATE: 1997-11-06
; PRIOR APPLICATION NUMBER: 08/531,842
; PRIOR FILING DATE: 1995-09-21
; PRIOR APPLICATION NUMBER: 08/275,196
; PRIOR FILING DATE: 1994-07-14
; PRIOR APPLICATION NUMBER: 07/976,805
; PRIOR FILING DATE: 1992-11-16
; PRIOR APPLICATION NUMBER: 07/453,445
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: 07/416,218
; PRIOR FILING DATE: 1989-10-02
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: VARIANT
; LOCATION: (1)...(50)
; OTHER INFORMATION: Xaa - Any Amino Acid
US-09-853-844-4

Query Match 50.5%; Score 53; DB 10; Length 50;
Best Local Similarity 55.0%; Pred. No. 0.02;
Matches 11; Conservative 3; Mismatches 6; Indels 6; Gaps 0;

Qy 2 LVEQASTSQASLYQIYNFDN 21
:|||||||:|:
Db 2 IVEQCCTSIQSLYQLENYCN 21

RESULT 7
US-10-028-410-3
; Sequence 3, Application US/10028410
; Patent No. US20020160955A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquié, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1-1
; CURRENT APPLICATION NUMBER: US/10/028,410
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US/09/477,924
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-410-3

Query Match 50.5%; Score 53; DB 9; Length 51;
Best Local Similarity 55.0%; Pred. No. 0.021;
Matches 11; Conservative 3; Mismatches 6; Indels 6; Gaps 0;

Qy 2 LVEQASTSQASLYQIYNFDN 21
:|||||||:|:
Db 32 IVEQCCTSIQSLYQLENYCN 51

RESULT 8
US-10-054-873-5
; Sequence 5, Application US/10054873
; Patent No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-054-873-5

Query Match 50.5%; Score 53; DB 9; Length 52;
Best Local Similarity 55.0%; Pred. No. 0.021;
Matches 11; Conservative 3; Mismatches 6; Indels 6; Gaps 0;

Qy 2 LVEQASTSQASLYQIYNFDN 21
:|||||||:|:
Db 33 IVEQCCTSIQSLYQLENYCN 52

RESULT 9
US-09-815-229-13
; Sequence 13, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDER
; FILE REFERENCE: P1786RUS
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17

```
; SEQ ID NO 13
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Portion of Insulin variant from EP 171,147.
US-09-815-229-13

Query Match          50.5%; Score 53; DB 10; Length 54;
Best Local Similarity 55.0%; Pred. No. 0.022;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
   :||| || ||||: | :
Db 35 IVEQCCTSCSLYLENYCN 54

RESULT 10
US-09-878-380-1
; Sequence 1, Application US/09878380
; Patent No. US20020160435A1
; GENERAL INFORMATION:
; APPLICANT: Fujirebio Inc.
; APPLICANT: KITAJIMA, Sachiko
; APPLICANT: KURANO, Yoshihiro
; APPLICANT: NAKATSUBO, Kaoru
; APPLICANT: NISHIZONO, Isao
; TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor
; FILE REFERENCE: 0760-0291P
; CURRENT APPLICATION NUMBER: US/09/878,380
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: JP 2000-174691
; PRIOR FILING DATE: 2000-06-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-878-380-1

Query Match          50.5%; Score 53; DB 9; Length 86;
Best Local Similarity 55.0%; Pred. No. 0.037;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
   :||| || ||||: | :
Db 67 IVEQCCTSCSLYLENYCN 86

RESULT 11
US-10-028-410-2
; Sequence 2, Application US/10028410
; Patent No. US20020160955A1
; GENERAL INFORMATION:
; APPLICANT: Dubaque, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1-1
; CURRENT APPLICATION NUMBER: US/10/028,410
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US/09/477,924
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 2
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-410-2

Query Match          50.5%; Score 53; DB 9; Length 86;
Best Local Similarity 55.0%; Pred. No. 0.037;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 2 LVEQASTSQASLYQIYNFDN 21
   :||| || ||||: | :
Db 67 IVEQCCTSCSLYLENYCN 86

RESULT 12
US-10-054-873-4
; Sequence 4, Application US/10054873
; Patent No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-0001300S
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 86 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-054-873-4

Query Match          50.5%; Score 53; DB 9; Length 86;
Best Local Similarity 55.0%; Pred. No. 0.037;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
   :||| || ||||: | :
Db 67 IVEQCCTSCSLYLENYCN 86

RESULT 13
US-09-947-563-4
; Sequence 4, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining insulin precursors having correctly bonded cystine
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrabow, Garrett & Dunner
; STREET: 1300 I Street, N.W.
```

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/947,563
APPLICATION NUMBER: US/09/947,563
FILING DATE: 07-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/134,836
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Leslie McDonnell
REGISTRATION NUMBER: 34,872
REFERENCE/DOCKET NUMBER: 02481.1600-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Protein
LOCATION: 1..96
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-947-563-4
Query Match 50.5%; Score 53; DB 9; Length 96;
Best Local Similarity 55.0%; Pred. No. 0.042;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
Qy 2 LIVEQASTSQASLYQIYNFDN 21
Db 77 IVEQCCTSCSLYQLENYCN 96
RESULT 14
US-10-054-873-6
; Sequence 6, Application US/10054873
; Patent No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/10/054,873
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/CN98/00052
FILING DATE: 31-MAR-1998
APPLICATION NUMBER: US 09/423,100
FILING DATE: 11-DEC-2000
ATTORNEY/AGENT INFORMATION:
NAME: Mycroft, Frank J
REGISTRATION NUMBER: 46,946
REFERENCE/DOCKET NUMBER: 020167-000130US
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 107 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-054-873-6
Query Match 50.5%; Score 53; DB 9; Length 107;
Best Local Similarity 55.0%; Pred. No. 0.048;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
Qy 2 LIVEQASTSQASLYQIYNFDN 21
Db 88 IVEQCCTSCSLYQLENYCN 107
RESULT 15
US-09-804-409A-9
; Sequence 9, Application US/09804409A
; Patent No. US20020155100A1
; GENERAL INFORMATION:
; APPLICANT: KIEFFER, TIMOTHY J.
; APPLICANT: CHEUNG, ANTHONY T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATED PROTEIN
; FILE REFERENCE: 029996/027 8721
; CURRENT APPLICATION NUMBER: US/09/804,409A
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 9
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-804-409A-9
Query Match 50.5%; Score 53; DB 9; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.049;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
Qy 2 LIVEQASTSQASLYQIYNFDN 21
Db 91 IVEQCCTSCSLYQLENYCN 110
RESULT 16
US-09-205-658-125
; Sequence 125, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080

;; EARLIER FILING DATE: 1998-05-15
;; NUMBER OF SEQ ID NOS: 328
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 125
;; LENGTH: 110
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-205-658-125

Query Match 50.5%; Score 53; DB 10; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.049; Mismatches 3; Indels 0; Gaps 0;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 2 LVEQASTSQASLYQIYNFDN 21
:|||||||
Db 91 IVEQCCTSCSLYLENYCN 110

RESULT 17
US-09-815-229-3
; Sequence 3, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
; FILE REFERENCE: P1786RIUS
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 3
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-229-3

Query Match 50.5%; Score 53; DB 10; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.049; Mismatches 3; Indels 0; Gaps 0;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 2 LVEQASTSQASLYQIYNFDN 21
:|||||||
Db 91 IVEQCCTSCSLYLENYCN 110

RESULT 18
US-09-280-030-63
; Sequence 63, Application US/09280030A
; Patent No. US20010021515A1
; GENERAL INFORMATION:
; APPLICANT: Sato, Seiji
; APPLICANT: Higashikuni, Naohiko
; APPLICANT: Kudo, Toshiyuki
; APPLICANT: Kondo, Masaaki
; TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR
; TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE
; FILE REFERENCE: 382.1026
; CURRENT APPLICATION NUMBER: US/09/280,030A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: JP10-87339/1998
; EARLIER FILING DATE: 1998-03-31
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 63
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Designated is
; OTHER INFORMATION: an amino acid sequence of

;; OTHER INFORMATION: MWFPsp-MWpmp10-Met-Proinsulin
US-09-280-030-63

Query Match 50.5%; Score 53; DB 10; Length 117;
Best Local Similarity 55.0%; Pred. No. 0.053;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 2 LVEQASTSQASLYQIYNFDN 21
:|||||||
Db 98 IVEQCCTSCSLYLENYCN 117

RESULT 19
US-09-736-611-12
; Sequence 12, Application US/09736611
; Patent No. US20010023069A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas
; APPLICANT: Ludvigsen, Svend
; APPLICANT: Kaarsholm, Niels
; TITLE OF INVENTION: Method For Making Insulin Precursors and
; TITLE OF INVENTION: Insulin Precursor Analogs
; FILE REFERENCE: 6058.200-US
; CURRENT APPLICATION NUMBER: US/09/736,611
; CURRENT FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/181,443
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 60/211,441
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: PA 1999 01868
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PA 2000 00440
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 124
; TYPE: PRT
; ORGANISM: N-terminal extension
US-09-736-611-12

Query Match 50.5%; Score 53; DB 10; Length 124;
Best Local Similarity 55.0%; Pred. No. 0.057;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 2 LVEQASTSQASLYQIYNFDN 21
:|||||||
Db 105 IVEQCCTSCSLYLENYCN 124

RESULT 20
US-09-740-359-12
; Sequence 12, Application US/09740359
; Patent No. US20010041787A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: Yield in yeast
; FILE REFERENCE: 6148.200-US
; CURRENT APPLICATION NUMBER: US/09/740,359
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 12
; LENGTH: 124
; TYPE: PRT
; ORGANISM: TA57 leader fused with N-terminally extended.....
US-09-740-359-12

Query Match      50.5%; Score 53; DB 10; Length 124;
Best Local Similarity 55.0%; Pred. No. 0.057;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY  2 LVEQASTSQASLYQIYNFDN 21
    :||| || ||||: | : |
Db  105 IVEQCCTSCSLYQLENYCN 124

RESULT 21
US-09-894-711-12
; Sequence 12, Application US/09894711
; Patent No. US20020137144A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: Insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: yield in yeast
; FILE REFERENCE: 6148.400-US
; CURRENT APPLICATION NUMBER: US/09/894,711
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 09/740,359
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-894-711-12

Query Match      50.5%; Score 53; DB 10; Length 124;
Best Local Similarity 55.0%; Pred. No. 0.057;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY  2 LVEQASTSQASLYQIYNFDN 21
    :||| || ||||: | : |
Db  105 IVEQCCTSCSLYQLENYCN 124

RESULT 22
US-09-894-711-18
; Sequence 18, Application US/09894711
; Patent No. US20020137144A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: Insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: yield in yeast
; FILE REFERENCE: 6148.400-US
; CURRENT APPLICATION NUMBER: US/09/894,711
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
```

```
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 09/740,359
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-894-711-18

Query Match      50.5%; Score 53; DB 10; Length 124;
Best Local Similarity 55.0%; Pred. No. 0.057;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY  2 LVEQASTSQASLYQIYNFDN 21
    :||| || ||||: | : |
Db  105 IVEQCCTSCSLYQLENYCN 124

RESULT 23
US-09-736-611-10
; Sequence 10, Application US/09736611
; Patent No. US20010023069A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas
; APPLICANT: Ludvigsen, Svend
; APPLICANT: Kaarsholm, Niels
; TITLE OF INVENTION: Method For Making Insulin Precursors and
; TITLE OF INVENTION: Insulin Precursor Analogs
; FILE REFERENCE: 6058.200-US
; CURRENT APPLICATION NUMBER: US/09/736,611
; CURRENT FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/181,443
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 60/211,441
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: PA 1999 01868
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: PA 2000 00440
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 125
; TYPE: PRT
; ORGANISM: N-terminal extension
US-09-736-611-10

Query Match      50.5%; Score 53; DB 10; Length 125;
Best Local Similarity 55.0%; Pred. No. 0.057;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY  2 LVEQASTSQASLYQIYNFDN 21
    :||| || ||||: | : |
Db  106 IVEQCCTSCSLYQLENYCN 125

RESULT 24
US-09-740-359-10
; Sequence 10, Application US/09740359
; Patent No. US20010041787A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: Insulin precursor analogues having improved fermentation
```

```
; TITLE OF INVENTION: yield in yeast
; FILE REFERENCE: 6148.200-US
; CURRENT APPLICATION NUMBER: US/09/740,359
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 125
; TYPE: PRT
; ORGANISM: TA39 leader fused with N-terminally extended
US-09-740-359-10
```

```
Query Match 50.5%; Score 53; DB 10; Length 125;
Best Local Similarity 55.0%; Pred. No. 0.057;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 2 LVEQASTSQASLYQIYNFDN 21
:||||| ||||| :|
Db 106 IVEQCCTSIQSLYQLENYCN 125
```

```
RESULT 25
US-09-894-711-10
; Sequence 10, Application US/09894711
; Patent No. US20020137144A1
; GENERAL INFORMATION:
; APPLICANT: Kjeldsen, Thomas Borglum
; APPLICANT: Ludvigsen, Svend
; TITLE OF INVENTION: Method for making insulin precursors and
; TITLE OF INVENTION: insulin precursor analogues having improved fermentation
; TITLE OF INVENTION: yield in yeast
; FILE REFERENCE: 6148.400-US
; CURRENT APPLICATION NUMBER: US/09/894,711
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PA 2000 00443
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: PA 1999 01869
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 60/211,081
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/181,450
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 09/740,359
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-894-711-10
```

```
Query Match 50.5%; Score 53; DB 10; Length 125;
Best Local Similarity 55.0%; Pred. No. 0.057;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 2 LVEQASTSQASLYQIYNFDN 21
:||||| ||||| :|
Db 106 IVEQCCTSIQSLYQLENYCN 125
```

```
Search completed: December 23, 2002, 07:20:18
Job time : 57.6471 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:17 ; Search time 10.7059 Seconds
(without alignments)
188.571 Million cell updates/sec

Title: US-09-574-443-7_COPY_1_21
Perfect score: 105
Sequence: 1 NLVEQASTSQASLYQIYNFDN 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	50.5	51	1 INWHP	insulin - sperm wh
2	53	50.5	51	1 INWHP	insulin - finback
3	53	50.5	84	1 IPDG	insulin precursor
4	53	50.5	96	2 PC7082	epidermal growth f
5	53	50.5	110	1 IPHU	insulin precursor
6	53	50.5	110	1 INRB	insulin precursor
7	53	50.5	110	1 IPDG	insulin precursor
8	53	50.5	110	2 B42179	insulin precursor
9	53	50.5	110	2 A42179	insulin precursor
10	53	50.5	110	2 JQ0178	insulin precursor
11	50	47.6	39	2 I49418	insulin I precursor
12	50	47.6	41	2 I49419	insulin II precursor
13	50	47.6	51	1 INWHP	insulin - sei whal
14	50	47.6	51	1 INEL	insulin - elephant
15	50	47.6	51	1 INHY	insulin - hamster
16	50	47.6	51	1 INMSSP	insulin - Egyptian
17	50	47.6	51	1 INKSO	insulin - common s
18	50	47.6	108	1 INMS1	insulin 1 precursor
19	50	47.6	108	2 A39883	insulin precursor
20	50	47.6	110	1 IPRT1	insulin 1 precursor
21	50	47.6	110	1 IPRT2	insulin 1 precursor
22	50	47.6	110	1 INMS2	insulin 2 precursor
23	50	47.6	110	2 I48166	insulin precursor
24	49	46.7	51	1 INKMA	insulin - Arabian
25	49	46.7	51	2 A59151	insulin precursor
26	49	46.7	86	1 IPHO	insulin precursor
27	49	46.7	105	1 IPBO	insulin precursor
28	48	45.7	51	1 INPO	insulin - crested
29	47	44.8	51	1 INCB	insulin - Chinchil

ALIGNMENTS

RESULT 1

INWHP
insulin - sperm whale
C:Species: Physeter catodon (sperm whale)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A93142; A90082
R:Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.
Nature 181, 1468-1469, 1958
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale
A:Reference number: A93142
A:Accession: A93142
A:Molecule type: protein
A:Residues: 1-30;31-51<ISH>
R:Harris, J.I.; Sanger, F.; Naughton, M.A.
Arch. Biochem. Biophys. 65, 427-428, 1956
A:Title: Species differences in insulin.
A:Reference number: A90082
A:Accession: A90082
A:Molecule type: protein
A:Residues: 1-30;31-51<HAR>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Dissulfide bonds: #status predicted

Query Match 50.5%; Score 53; DB 1; Length 51;
Best Local Similarity 55.0%; Pred. No. 0.071;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
DB 32 IVEQCCTSGICSLYLENYCN 51

RESULT 2

INWHP
insulin - finback whale (tentative sequence)
C:Species: Balaenoptera physalus (finback whale, common rorqual)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A91918
R:Hama, H.; Titani, K.; Sakaki, S.; Narita, K.
J. Biochem. 56, 285-293, 1964
A:Title: The amino acid sequence in fin-whale insulin.
A:Reference number: A91918
A:Accession: A91918
A:Molecule type: protein
A:Residues: 1-30;31-51<HAM>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>

F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 50.5%; Score 53; DB 1; Length 51;
Best Local Similarity 55.0%; Pred. No. 0.071;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:111 111111: 1:
Db 32 IVEQCCTSIQSLYLENYCN 51

RESULT 3

IPPG

insulin precursor - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 16-Jul-1999
C;Accession: A01583; A94572; S16492; A60835; B60835
R;Chance, R.E.; Ellis, R.M.; Bromer, W.W.

A;Title: Porcine proinsulin: characterization and amino acid sequence.
A;Reference number: A94240; MUID:68286485; PMID:5657063

A;Accession: A01583
A;Molecule type: protein
A;Residues: 1-34,'Q',36-84 <CHA>
R;Chance, R.E.

submitted to the Atlas, July 1970
A;Reference number: A94572

A;Accession: A94572
A;Molecule type: protein

A;Residues: 1-84 <CH2>
R;Brown, H.; Sanger, F.; Kitai, R.

Biochem. J. 60, 556-565, 1955
A;Title: The structure of pig and sheep insulins.

A;Reference number: A90344
A;Accession: S16492

A;Molecule type: protein
A;Residues: 1-30,31-51 <BRO>

R;Snel, L.; Damgaard, U.
Horm. Metab. Res. 20, 476-480, 1988

A;Title: Proinsulin heterogeneity in pigs.

A;Reference number: A60835; MUID:89032178; PMID:3181865

A;Accession: A60835

A;Molecule type: protein

A;Residues: 33-38,40-62 <SNE>

A;Note: the authors report the characterization of a connecting peptide variant lacking

A;Accession: B60835

A;Molecule type: protein

A;Residues: 33-62 <SN2>

R;Blundell, T.; Dodson, G.; Hodgkin, D.; Mercola, D.

Adv. Protein Chem. 26, 279-402, 1972

A;Title: Insulin. the structure in the crystal and its reflection in chemistry and biology

A;Reference number: A90017

A;Contents: annotation; X-ray crystallography, 1.9 angstroms

C;Superfamily: insulin

C;Keywords: hormone; pancreas

F;1-30/Domain: insulin chain B #status experimental <BCH>

F;33-63/Domain: connecting peptide #status experimental <CPEP>

F;64-84/Domain: insulin chain A #status experimental <ACH>

F;7-70,19-83,69-74/Disulfide bonds: #status experimental

Query Match 50.5%; Score 53; DB 1; Length 84;
Best Local Similarity 55.0%; Pred. No. 0.13;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:111 111111: 1:
Db 65 IVEQCCTSIQSLYLENYCN 84

RESULT 4

PC7082

epidermal growth factor/single chain insulin fusion protein - Bacillus brevis (f)
C;Species: Bacillus brevis
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C;Accession: PC7082; PC7083

R;Koh, M.; Hanagata, H.; Ebisu, S.; Morihara, K.; Takagi, H.
Biosci. Biotechnol. Biochem. 64, 1079-1081, 2000

A;Title: Use of Bacillus brevis for synthesis and secretion of Des-B30 single-chain

A;Reference number: PC7082; MUID:20335834; PMID:10879487

A;Accession: PC7082

A;Molecule type: DNA

A;Residues: 1-96 <KOH>

A;Accession: PC7083

A;Molecule type: protein

A;Residues: 19-28 <KO2>

C;Genetics:

A;Gene: egf-sci

C;Superfamily: insulin

C;Keywords: fusion protein

Query Match 50.5%; Score 53; DB 2; Length 96;

Best Local Similarity 55.0%; Pred. No. 0.15;

Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21

:111 111111: 1:
Db 77 IVEQCCTSIQSLYLENYCN 96

RESULT 5

IPHU

insulin precursor [validated] - human

N;Alternate names: preproinsulin

C;Species: Homo sapiens (man)

C;Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 08-Dec-2000

C;Accession: A93222; A94253; A93216; A94251; A93144; A91186; I58114; A01

R;Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.

Nature 284, 26-32, 1980

A;Title: Sequence of the human insulin gene.

A;Reference number: A93222; MUID:80120725; PMID:6243748

A;Accession: A93222

A;Molecule type: DNA

A;Residues: 1-110 <BEL>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Ullrich, A.; Dull, T.J.; Gray, A.; Broslus, J.; Sures, I.

Science 209, 612-615, 1980

A;Title: Genetic variation in the human insulin gene.

A;Reference number: A94253; MUID:80236313; PMID:6248962

A;Accession: A94253

A;Molecule type: DNA

A;Residues: 1-110 <ULL>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.

Nature 282, 525-527, 1979

A;Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.

A;Reference number: A93216; MUID:80054779; PMID:503234

A;Accession: A93216

A;Molecule type: mRNA

A;Residues: 1-110 <BEL2>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.

Science 208, 57-59, 1980

A;Title: Nucleotide sequence of human preproinsulin complementary DNA.

A;Reference number: A94251; MUID:80147417; PMID:6927840

A;Accession: A94251

A;Molecule type: mRNA

A;Residues: 1-110 <SUR>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Nicoll, D.S.H.W.; Smith, L.F.

Nature 187, 483-485, 1960

A;Title: Amino-acid sequence of human insulin.

A;Reference number: A93144

A;Accession: A93144

A:Molecule type: protein
A:Residues: 25-54;90-110 <NIG>
R:Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.
J. Biol. Chem. 246, 1375-1386, 1971
A:Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pan
A:Reference number: A92075; MUID:71116410; PMID:5101771
A:Molecule type: protein
A:Residues: 57-87 <OYE>
R:Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
Eur. J. Biochem. 20, 190-199, 1971
A:Title: Amino acid sequence of the C-peptide of human proinsulin.
A:Reference number: A91186; MUID:71257722; PMID:5560404
A:Accession: A91186
A:Molecule type: protein
A:Residues: 57-87 <KOA>
R:Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boltard, C.; Froguet, P.; Lathrop, M.; Bell
Nature Genet. 4, 305-310, 1993
A:Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment
A:Reference number: I58114; MUID:93364428; PMID:8358440
A:Accession: I58114
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-59,63-110 <RES>
A:Cross-references: GB:I15440; NID:g307071; PIDN:AAA59179.1; PID:g307072
R:Stieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.
Helv. Chim. Acta 57, 2617-2621, 1974
A:Title: Totalsynthese von humaninsulin unter gezielter Bildung der Disulfidbindungen.
A:Reference number: A91636; MUID:75077277; PMID:4443293
A:Contents: annotation; synthesis
A:Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical
A:Note: article in German with English abstract
R:Naithani, V.K.
Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
A:Title: The synthesis of C-peptide of human proinsulin.
A:Reference number: A91658; MUID:75040007; PMID:4803504
A:Contents: annotation; synthesis of residues 57-87
R:Geiger, R.; Jaeger, G.; Koenig, W.
Chem. Ber. 106, 2347-2352, 1973
A:Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9
A:Reference number: A90914
A:Contents: annotation; synthesis of residues 57-87
R:Kaufmann, J.E.; Irmlinger, J.C.; Halban, P.A.
Biochem. J. 310, 869-874, 1995
A:Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide junction
A:Reference number: S58661; MUID:96013185; PMID:7575420
A:Contents: annotation; site-directed mutagenesis study of proteolytic processing
C:Genetics:
A:Gene: GDB:INS
A:Cross-references: GDB:I19349; OMIM:176730
A:Map position: 11p15.5-11p15.5
A:Introns: 63/1
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54/Domain: insulin chain B #status experimental <BCH>
F:25-54,90-110/Product: insulin #status experimental <MAT>
F:57-87/Domain: connecting C peptide #status experimental <CPEP>
F:90-110/Domain: insulin chain A #status experimental <ACH>
F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 50.5%; Score 53; DB 1; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.17;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 2 LVEQASTSQASLYQYINFDN 21
:|||||:|
Db 91 IVEQCCTSCSLYQLENYCN 110

RESULT 6

INRB

Insulin precursor - rabbit

N:Alternate names: preproinsulin
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 24-Apr-1984 #sequence_revision 23-Aug-1997 #text_change 18-Jun-1999
C:Accession: A53438; A01581
R:Devaskar, S.U.; Giddings, S.J.; Rajakumar, P.A.; Carnaghi, L.R.; Menon, R.K.; Z
J. Biol. Chem. 269, 8445-8454, 1994
A:Title: Insulin gene expression and insulin synthesis in mammalian neuronal cell
A:Reference number: A53438; MUID:94179230; PMID:8132571
A:Accession: A53438
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-110 <DEV>
A:Cross-references: GB:U03610; NID:g467970; PIDN:AAA19033.1; PID:g467971
R:Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A:Title: Species variation in the amino acid sequence of insulin.
A:Reference number: A90029; MUID:66160119; PMID:5949593
A:Accession: A01581
A:Molecule type: protein
A:Residues: 25-54;90-110 <SMI>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54/Domain: insulin chain B #status experimental <BCH>
F:25-54,90-110/Product: insulin #status experimental <MAT>
F:57-87/Domain: connecting C peptide #status predicted <CPEP>
F:90-110/Domain: insulin chain A #status experimental <ACH>
F:31-96,43-109,95-100/Disulfide bonds: #status predicted
Query Match 50.5%; Score 53; DB 1; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.17;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 2 LVEQASTSQASLYQYINFDN 21

Db 91 IVEQCCTSCSLYQLENYCN 110

RESULT 7

IPDG

Insulin precursor - dog

C:Species: Canis lupus familiaris (dog)

C:Date: 24-Apr-1984 #sequence_revision 15-Nov-1984 #text_change 16-Jul-1999

C:Accession: A92413; A01587; S16493

R:Kwok, S.C.M.; Chan, S.J.; Steiner, D.F.

J. Biol. Chem. 258, 2357-2363, 1983

A:Title: Cloning and nucleotide sequence analysis of the dog insulin gene. Coded

A:Reference number: A92413; MUID:83109071; PMID:6296142

A:Accession: A92413

A:Molecule type: DNA

A:Residues: 1-110 <SMI>

A:Cross-references: GB:V00179; GB:J00042; NID:g994; PIDN:CAA23475.1; PID:g995

R:Smith, L.F.

Am. J. Med. 40, 662-666, 1966

A:Title: Species variation in the amino acid sequence of insulin.

A:Reference number: A90029; MUID:66160119; PMID:5949593

A:Accession: A01587

A:Molecule type: protein

A:Residues: 25-54;90-110 <SMIT>

R:Peterson, J.D.; Nehrlisch, S.; Oyer, P.E.; Steiner, D.F.

J. Biol. Chem. 247, 4866-4871, 1972

A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog I

A:Reference number: A92111; MUID:72258016; PMID:4626369

A:Accession: S16493

A:Molecule type: protein

A:Residues: 65-85, 1, 87 <PET>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-54/Domain: insulin chain B #status experimental <BCH>

F:25-54,90-110/Product: insulin #status experimental <MAT>

F:57-87/Domain: connecting peptide #status predicted <CPEP>

F:90-110/Domain: insulin chain A #status experimental <ACH>

F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 50.5%; Score 53; DB 1; Length 110;
 Best Local Similarity 55.0%; Pred. No. 0.17;
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
 :||| || ||||: |
 Db 91 IVEQCCTSIQSILYQLENYCN 110

RESULT 8
 Insulin precursor - green monkey
 C:Species: Cercopithecus aethiops (green monkey, grivet)
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
 C:Accession: B42179; A05232; S16494; S22056
 R:Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate of
 A:Reference number: A42179; MUID:92219953; PMID:1560757
 A:Accession: B42179
 A:Molecule type: DNA
 A:Residues: 1-110 <SEI>
 A:Cross-references: EMBL:X61092; NID:922808; PIDN:CAA43403.1; PID:g22809
 A:Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBIP:95194)
 R:Peterson, J.D.; Nehrlich, S.; Over, P.E.; Steiner, D.F.
 J. Biol. Chem. 247, 4866-4871, 1972
 A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin
 A:Reference number: A92111; MUID:72258016; PMID:4626369
 A:Accession: A05232
 A:Molecule type: protein
 A:Residues: 57-87 <PET>
 C:Genetics:
 A:Introns: 63/1
 C:Superfamily: insulin
 C:Keywords: hormone; pancreas
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54/Domain: insulin chain B #status predicted <BCH>
 F:55-89/Domain: insulin connecting peptide #status predicted <CPEP>
 F:90-110/Domain: insulin chain A #status predicted <ACH>
 F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 50.5%; Score 53; DB 2; Length 110;
 Best Local Similarity 55.0%; Pred. No. 0.17;
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
 :||| || ||||: |
 Db 91 IVEQCCTSIQSILYQLENYCN 110

RESULT 9
 Insulin precursor - chimpanzee
 C:Species: Pan troglodytes (chimpanzee)
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
 C:Accession: A42179; S22058
 R:Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate of
 A:Reference number: A42179; MUID:92219953; PMID:1560757
 A:Accession: A42179
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-110 <SEI>
 A:Cross-references: EMBL:X61089; NID:g38251; PIDN:CAA43403.1; PID:g38252
 A:Note: sequence extracted from NCBI backbone (NCBIP:95067)
 C:Genetics:
 A:Introns: 63/1
 C:Superfamily: insulin

Query Match 50.5%; Score 53; DB 2; Length 110;
 Best Local Similarity 55.0%; Pred. No. 0.17;
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
 :||| || ||||: |
 Db 91 IVEQCCTSIQSILYQLENYCN 110

RESULT 10
 Insulin precursor - crab-eating macaque
 C:Species: Macaca fascicularis (crab-eating macaque)
 C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 16-Jul-1999
 C:Accession: JQ0178
 R:Wetkham, W.; Groneberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.
 Gene 19, 179-183, 1982
 A:Title: The nucleotide sequence of cDNA coding for preproinsulin from the primate
 A:Reference number: JQ0178; MUID:83080474; PMID:6184262
 A:Accession: JQ0178
 A:Molecule type: mRNA
 A:Residues: 1-110 <WET>
 A:Cross-references: GB:J00336; NID:g342121; PIDN:AAA36849.1; PID:g342122
 C:Superfamily: insulin
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54,90-110/Product: insulin #status predicted <WAT>
 F:55-89/Domain: insulin chain B #status predicted <BCH>
 F:90-110/Domain: insulin connecting C peptide #status predicted <CPT>
 F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 50.5%; Score 53; DB 2; Length 110;
 Best Local Similarity 55.0%; Pred. No. 0.17;
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
 :||| || ||||: |
 Db 91 IVEQCCTSIQSILYQLENYCN 110

RESULT 11
 Insulin I precursor - western wild mouse (fragment)
 C:Species: Mus spretus (western wild mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999
 C:Accession: I49418
 R:Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadea
 Mamm. Genome 5, 349-355, 1994
 A:Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.
 A:Reference number: I48934; MUID:94319082; PMID:8043949
 A:Accession: I49418
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-39 <RES>
 A:Cross-references: EMBL:U05729; NID:g497062; PIDN:AAB60473.1; PID:g497063
 C:Superfamily: insulin

Query Match 47.6%; Score 50; DB 2; Length 39;
 Best Local Similarity 50.0%; Pred. No. 0.16;
 Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
 :||| || ||||: |
 Db 20 IVEQCCTSIQSILYQLENYCN 39

RESULT 12
 Insulin II precursor - western wild mouse (fragment)
 C:Species: Mus spretus (western wild mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999
 C:Accession: I49419
 R:Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadea

Mamm. Genome 5, 349-355, 1994

A:Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.

A:Reference number: I48934; MUID:94319082; PMID:8043949

A:Accession: I49419

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-41 <RES>

A:CROSS-references: EMBL:U05730; NID:g497064; PIDN:AAB60474.1; PID:g497065

C:Superfamily: insulin

Query Match 47.6%; Score 50; DB 2; Length 41;

Best Local Similarity 50.0%; Pred. No. 0.17;

Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21

DB 22 IVQQCTSCSLYQLENYCN 41

RESULT 13

INWHIS

Insulin - sei whale

C:Species: Balanoptera borealis (sei whale)

C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999

C:Accession: A01582

R: Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.

Nature 181, 1468-1469, 1958

A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin.

A:Reference number: A93142

A:Accession: A01582

A:Molecule type: protein

A:Residues: 1-30;31-51 <ISH>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:1-30,31-51/Product: insulin #status experimental <MAT>

F:31-51/Domain: insulin chain A #status experimental <ACH>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 47.6%; Score 50; DB 1; Length 51;

Best Local Similarity 50.0%; Pred. No. 0.22;

Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21

DB 32 IVEQCCASTCSLYQLENYCN 51

RESULT 14

INEL

Insulin - elephant

C:Species: Elephantidae gen. sp. (elephant)

C:Date: 24-Apr-1984 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999

C:Accession: A01584

R: Smith, L.F.

Am. J. Med. 40, 662-666, 1966

A:Title: Species variation in the amino acid sequence of insulin.

A:Reference number: A90029; MUID:66160119; PMID:5949593

A:Accession: A01584

A:Molecule type: protein

A:Residues: 1-30;31-51 <SMI>

A:Note: the species of elephant is not given, but it is most probably the Indian elephant

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:1-30,31-51/Product: insulin #status experimental <MAT>

F:31-51/Domain: insulin chain A #status experimental <ACH>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 47.6%; Score 50; DB 1; Length 51;

Best Local Similarity 50.0%; Pred. No. 0.22;

Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21

DB 32 IVEQCCCTGVCSLYQLENYCN 51

RESULT 15

INHJ

Insulin - hamster

C:Species: Cricetinae gen. sp. (hamster)

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999

C:Accession: A91456

R: Neelon, F.A.; Delcher, H.K.; Steinman, H.; Lebovitz, H.E.

Fed. Proc. 32, 300, 1973

A:Title: Structure of hamster insulin: comparison with a tumor insulin.

A:Reference number: A91456

A:Accession: A91456

A:Molecule type: protein

A:Residues: 1-30;31-51 <NEE>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:1-30,31-51/Product: insulin #status experimental <MAT>

F:31-51/Domain: insulin chain A #status experimental <ACH>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 47.6%; Score 50; DB 1; Length 51;

Best Local Similarity 50.0%; Pred. No. 0.22;

Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21

DB 32 IVQQCTSCSLYQLENYCN 51

RESULT 16

INMSSP

Insulin - Egyptian spiny mouse (tentative sequence)

C:Species: Acomys cahirinus (Egyptian spiny mouse)

C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000

C:Accession: A01591

R: Buenzli, H.F.; Humbel, R.E.

Hoppe-Seyler's Z. Physiol. Chem. 353, 444-450, 1972

A:Title: Isolation and partial structural analysis of insulin from mouse (Mus mu

A:Reference number: A01591; MUID:72189454; PMID:5028210

A:Contents: composition

A:Accession: A01591

A:Molecule type: protein

A:Residues: 1-30;31-51 <BUE>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status predicted <BCH>

F:1-30,31-51/Product: insulin #status predicted <MAT>

F:31-51/Domain: insulin chain A #status predicted <ACH>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 47.6%; Score 50; DB 1; Length 51;

Best Local Similarity 50.0%; Pred. No. 0.22;

Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21

DB 32 IVQQCTSCSLYQLENYCN 51

RESULT 17

INNSQ

Insulin - common squirrel monkey

C:Species: Saimiri sciureus (common squirrel monkey)

C:Date: 18-Oct-1991 #sequence_revision 02-May-1994 #text_change 16-Jul-1999

C:Accession: B39258; A39258

R: Yu, J.H.; Eng, J.; Yalow, R.S.

Proc. Natl. Acad. Sci. U.S.A. 87, 9766-9768, 1990

A:Title: Isolation and amino acid sequences of squirrel monkey (Saimiri sciurea)

A:Reference number: A39258; MUID:91088593; PMID:2263627

A:Accession: B39258

A:Molecule type: protein

A:Residues: 1-30 <YUB>

A:Accession: A39258

A:Molecule type: protein

A:Residues: 31-51 <YUA>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: Insulin chain B #status experimental <BCH>

F:1-30,31-51/Product: Insulin #status experimental <MAT>

F:31-51/Domain: Insulin chain A #status experimental <ACH>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 47.6%; Score 50; DB 1; Length 51;

Best Local Similarity 50.0%; Pred. No. 0.22;

Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVQASTSQASLYQIYNFDN 21

II:II II IIII:II

Db 32 VVQCTCSICSLYQLQNYCN 51

RESULT 18

INNS1

Insulin 1 precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 24-Apr-1984 #sequence_revision 14-Jul-1994 #text_change 18-Jun-1999

C:Accession: B26342; A48172; A01592; B61012

R:Wentworth, B.M.; Schaefer, I.M.; Villa-Komaroff, L.; Chirgwin, J.M.

J. Mol. Evol. 23, 305-312, 1986

A:Title: Characterization of the two nonallelic genes encoding mouse preproinsulin.

A:Reference number: A92965; MUID:87169768; PMID:3104603

A:Accession: B26342

A:Molecule type: DNA

A:Residues: 1-108 <VEN>

A:Cross-references: GB:X04725; NID:g52712; PIDN:CAA28434.1; PID:g52713

R:Sawa, T.; Ohgaku, S.; Morioka, H.; Yano, S.

J. Mol. Endocrinol. 5, 61-67, 1990

A:Title: Molecular cloning and DNA sequence analysis of preproinsulin genes in the NON

A:Reference number: A48172; MUID:90372989; PMID:2397023

A:Accession: A48172

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-108 <SAW>

R:Buenzli, H.F.; Glatthaar, B.; Kunz, P.; Muelhaupt, E.; Humbel, R.E.

Hoppe-Seyler's Z. Physiol. Chem. 353, 451-458, 1972

A:Title: Amino acid sequence of the two insulins from mouse (Mus musculus).

A:Reference number: A01592; MUID:72189455; PMID:5063718

A:Accession: A01592

A:Molecule type: protein

A:Residues: 25-54;88-108 <BUE>

R:Linde, S.; Nielsen, J.H.; Hansen, B.; Welinder, B.S.

J. Chromatogr. 462, 243-254, 1989

A:Title: Reversed-phase high-performance liquid chromatographic analyses of insulin bios

A:Reference number: A61012; MUID:89292078; PMID:2661585

A:Accession: B61012

A:Molecule type: protein

A:Residues: 57-85 <LIN>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-54/Domain: Insulin chain B #status experimental <BCH>

F:25-54,88-108/Product: Insulin #status experimental <MAT>

F:57-85/Domain: connecting peptide #status experimental <CPEP>

F:88-108/Domain: Insulin chain A #status experimental <ACH>

F:31-94,43-107,93-98/Disulfide bonds: #status predicted

Query Match 47.6%; Score 50; DB 1; Length 108;

Best Local Similarity 50.0%; Pred. No. 0.52;

Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVQASTSQASLYQIYNFDN 21

II:II II IIII:II

Db 89 IVQCTCSICSLYQLENYCN 108

RESULT 19

A39883

Insulin precursor - douroucouli

C:Species: Aotus trivirgatus (douroucouli, night monkey, owl monkey)

C:Date: 27-Nov-1991 #sequence_revision 27-Nov-1991 #text_change 16-Jul-1999

C:Accession: A39883

R:Seino, S.; Steiner, D.F.; Bell, G.I.

Proc. Natl. Acad. Sci. U.S.A. 84, 7423-7427, 1987

A:Title: Sequence of a New World primate insulin having low biological potency a

A:Reference number: A39883; MUID:88041119; PMID:3118367

A:Accession: A39883

A:Molecule type: DNA

A:Residues: 1-108 <SEI>

A:Cross-references: GB:J02989; NID:g176555; PIDN:AAA35374.1; PID:g176556

C:Superfamily: insulin

Query Match 47.6%; Score 50; DB 2; Length 108;

Best Local Similarity 50.0%; Pred. No. 0.52;

Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVQASTSQASLYQIYNFDN 21

II:II II IIII:II

Db 89 VVQCTCSICSLYQLQNYCN 108

RESULT 20

IPRT1

Insulin 1 precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 24-Sep-1999

C:Accession: A90788; A94231; B92120; I51945; A01589

R:Cordell, B.; Bell, G.; Tischer, E.; DeNoto, F.M.; Ullrich, A.; Pictet, R.; Rut

Cell 18, 533-543, 1979

A:Title: Isolation and characterization of a cloned rat insulin gene.

A:Reference number: A90788; MUID:80045034; PMID:498283

A:Accession: A90788

A:Molecule type: DNA

A:Residues: 1-110 <COR>

R:Lomedico, P.; Rosenthal, N.; Efstratiadis, A.; Gilbert, W.; Kolodner, R.; Tiza

Cell 18, 545-558, 1979

A:Title: The structure and evolution of the two nonallelic rat preproinsulin gen

A:Reference number: A90789; MUID:80045035; PMID:498284

A:Accession: A90789

A:Molecule type: DNA

A:Residues: 1-110 <LOM>

A:Cross-references: GB:J00747; NID:g204956; PIDN:AAA41442.1; PID:g204957

R:Steiner, D.F.; Clark, J.L.; Nolan, C.; Rubenstein, A.H.; Margoliash, E.; Aten,

Recent Prog. Horm. Res. 25, 207-282, 1969

A:Title: Proinsulin and the biosynthesis of insulin.

A:Reference number: A94231; MUID:70067613; PMID:4311938

A:Accession: A94231

A:Molecule type: protein

A:Residues: 25-54;90-110 <STE>

R:Tagger, H.S.; Steiner, D.F.

J. Biol. Chem. 247, 7936-7940, 1972

A:Title: Primary structures of the proinsulin connecting peptides of the rat and

A:Reference number: A92120; MUID:73061498; PMID:4640931

A:Accession: B92120

A:Molecule type: protein

A:Residues: 57-87 <TAG>

R:Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstratiadis, A.; Gilbert, W.

Ann. N. Y. Acad. Sci. 343, 425-432, 1980

A:Title: The structure of rat preproinsulin genes.

A:Reference number: I51945; MUID:80240379; PMID:6249167

A:Accession: I51945

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-110 <RES>
 A:Cross-references: GB:M25584; NID:g204947; PIDN:AAA41439.1; PID:g204948
 C:Genetics:
 A:Gene: INS1
 C:Superfamily: insulin

C:Keywords: hormone; pancreas
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54/Domain: insulin chain B #status experimental <BCH>
 F:25-54,90-110/Product: insulin #status experimental <MAT>
 F:57-87/Domain: connecting peptide #status experimental <CPEP>
 F:90-110/Domain: insulin chain A #status experimental <ACH>
 F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 47.6%; Score 50; DB 1; Length 110;
 Best Local Similarity 50.0%; Pred. No. 0.53;
 Matches 10; Conservative 4; Mismatches 6; Indels 6; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
 :|:| || ||||: |:
 Db 91 IVDQCTSCSLYQLENYCN 110

RESULT 21

Insulin 2 precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 24-Sep-1999
 C:Accession: B90789; B94231; C92120; I64880; A01590; B92120
 R:Lomedico, P.; Rosenthal, N.; Efstratiadis, A.; Gilbert, W.; Kolodner, R.; Tizard, R.
 Cell 18, 545-558, 1979
 A:Title: The structure and evolution of the two nonallelic rat preproinsulin genes.
 A:Reference number: A90789; MUID:80045035; PMID:498284
 A:Accession: B90789
 A:Molecule type: DNA

A:Residues: 1-110 <LON>
 A:Cross-references: GB:J00748; NID:g204958; PIDN:AAA41443.1; PID:g204959
 R:Steiner, D.F.; Clark, J.L.; Nolan, C.; Rubenstein, A.H.; Margoliash, E.; Aten, B.; Oye
 Recent Prog. Horm. Res. 25, 207-282, 1969
 A:Title: Proinsulin and the biosynthesis of insulin.
 A:Reference number: A94231; MUID:70067613; PMID:4311938
 A:Accession: B94231
 A:Molecule type: protein

A:Residues: 25-54;90-110 <STE>
 R:Tager, H.S.; Steiner, D.F.
 J. Biol. Chem. 247, 7936-7940, 1972
 A:Title: Primary structures of the proinsulin connecting peptides of the rat and horse.
 A:Reference number: A92120; MUID:73061498; PMID:4640931
 A:Accession: C92120
 A:Molecule type: protein

A:Residues: 57-87 <PAG>
 R:Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstratiadis, A.; Gilbert, W.
 Ann. N. Y. Acad. Sci. 343, 425-432, 1980
 A:Title: The structure of rat preproinsulin genes.
 A:Reference number: I51945; MUID:80240379; PMID:6249167
 A:Accession: I64880
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-110 <RES>
 A:Cross-references: GB:M25585; NID:g204950; PIDN:AAA41440.1; PID:g204952
 C:Genetics:
 A:Gene: INS2
 A:Introns: 63/1
 C:Superfamily: insulin

C:Keywords: hormone; pancreas
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54/Domain: insulin chain B #status experimental <BCH>
 F:25-54,90-110/Product: insulin #status experimental <MAT>
 F:57-87/Domain: connecting peptide #status experimental <CPEP>
 F:90-110/Domain: insulin chain A #status experimental <ACH>
 F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 47.6%; Score 50; DB 1; Length 110;
 Best Local Similarity 50.0%; Pred. No. 0.53;

Matches 10; Conservative 4; Mismatches 6; Indels 6; Gaps 0;
 QY 2 LVEQASTSQASLYQIYNFDN 21
 :|:| || ||||: |:
 Db 91 IVDQCTSCSLYQLENYCN 110

RESULT 22

Insulin 2 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 31-Mar-1992 #sequence_revision 14-Jul-1994 #text_change 18-Jun-1999
 C:Accession: A26342; B48172; A61012; B01592
 R:Wentworth, B.M.; Schaefer, I.M.; Villa-Komaroff, L.; Chirgwin, J.M.
 J. Mol. Evol. 23, 305-312, 1986
 A:Title: Characterization of the two nonallelic genes encoding mouse preproinsulin
 A:Reference number: A92965; MUID:87169768; PMID:3104603
 A:Accession: A26342
 A:Molecule type: DNA

A:Residues: 1-110 <WEN>
 A:Cross-references: GB:X04724; NID:g52714; PIDN:CAA28433.1; PID:g52715
 R:Sawa, T.; Ohgaku, S.; Morioka, H.; Yano, S.
 J. Mol. Endocrinol. 5, 61-67, 1990
 A:Title: Molecular cloning and DNA sequence analysis of preproinsulin genes in th
 A:Reference number: A48172; MUID:90372989; PMID:2397023
 A:Accession: B48172
 A:Status: not compared with conceptual translation

A:Molecule type: DNA
 A:Residues: 1-110 <SAW>
 R:Linde, S.; Nielsen, J.H.; Hansen, B.; Welinder, B.S.
 J. Chromatogr. 462, 243-254, 1989
 A:Title: Reversed-phase high-performance liquid chromatographic analyses of insul
 A:Reference number: A61012; MUID:89292078; PMID:2661585
 A:Accession: A61012
 A:Molecule type: protein

A:Residues: 57-87 <LIN>
 R:Buenzli, H.F.; Glatthaar, B.; Kunz, P.; Muelhaupt, E.; Humbel, R.E.
 Hoppe-Seyler's Z. Physiol. Chem. 353, 451-458, 1972
 A:Title: Amino acid sequence of the two insulins from mouse (Mus musculus).
 A:Reference number: A01592; MUID:72189455; PMID:5063718
 A:Accession: B01592
 A:Molecule type: protein

A:Residues: 25-54;90-110 <BUE>
 C:Genetics:
 A:Introns: 63/1
 C:Superfamily: insulin

C:Keywords: hormone; pancreas
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54/Domain: insulin chain B #status experimental <BCH>
 F:25-54,90-110/Product: insulin #status experimental <MAT>
 F:57-87/Domain: connecting peptide #status experimental <CPEP>
 F:90-110/Domain: insulin chain A #status experimental <ACH>
 F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 47.6%; Score 50; DB 1; Length 110;
 Best Local Similarity 50.0%; Pred. No. 0.53;

Matches 10; Conservative 4; Mismatches 6; Indels 6; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
 :|:| || ||||: |:
 Db 91 IVDQCTSCSLYQLENYCN 110

RESULT 23

Insulin precursor - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999
 R:Bell, G.I.; Sanchez-Pescador, R.
 Diabetes 33, 297-300, 1984
 A:Title: Sequence of a cDNA encoding Syrian hamster preproinsulin.
 A:Reference number: I48166; MUID:84133036; PMID:6365663

OY 2 LVEQASTSQASLYQIYNFDN 21
Db 32 IVEQCCASVCSLYQLENYCN 51
Search completed: December 23, 2002, 07:16:12
Job time : 11.7059 secs

A:Accession: I48166
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-110 <RES>
A:Cross-references: GB:M26328; NID:g191420; PIDN:AAA37089.1; PID:g305360
C:Superfamily: insulin

Query Match 47.6%; Score 50; DB 2; Length 110;
Best Local Similarity 50.0%; Pred. No. 0.53;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
Db 91 IVDQCTSGSLYQLENYCN 110

RESULT 24
INCMA
insulin - Arabian camel (tentative sequence)
C:Species: Camelus dromedarius (Arabian camel)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A92782
R:Danho, W.O.
J. Fac. Med. Baghdad 14, 16-28, 1972
A:Title: The isolation and characterization of insulin of camel (Camelus dromedarius).
A:Reference number: A92782

A:Accession: A92782
A:Molecule type: protein
A:Residues: 1-30;31-51 <DAN>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>
F:31-51/Product: Insulin #status experimental <MAT>
F:31-51/Domain: Insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 46.7%; Score 49; DB 1; Length 51;
Best Local Similarity 50.0%; Pred. No. 0.33;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
Db 32 IVEQCCASVCSLYQLENYCN 51

RESULT 25
A59151
insulin precursor - jack bean (fragments)
N:Alternate names: hypoglycemic agent; plant insulin
C:Species: Canavalia ensiformis (jack bean)
C:Date: 07-Dec-1999 #sequence_revision 07-Dec-1999 #text_change 10-Dec-1999
C:Accession: B59151
R:Oliveira, A.E.A.; Machado, O.L.T.; Gomes, V.M.; Xavier-Neto, J.; Pereira, A.C.P.; Vieira
Protein Pept. Lett. 6, 15-21, 1999
A:Title: Jack bean seed coat contains a protein with complete sequence homology to bovin
A:Reference number: A59151
A:Accession: B59151
A:Molecule type: protein
A:Residues: 1-30 <MACB>
A:Accession: A59151
A:Molecule type: protein
A:Residues: 31-51 <MACA>
C:Comment: The two chains are probably produced from the same precursor.
C:Superfamily: insulin
F:1-30,31-51/Product: Insulin #status experimental <MAT>
F:1-30/Domain: chain B #status experimental <CHB>
F:31-51/Domain: chain A #status experimental <CHA>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 46.7%; Score 49; DB 2; Length 51;
Best Local Similarity 50.0%; Pred. No. 0.33;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:17 ; Search time 6.17647 Seconds
(without alignments)
141.020 Million cell updates/sec

Title: US-09-574-443-7_COPY_1_21
Perfect score: 105
Sequence: 1 NLVEQASTSQASLYQIYNFDN 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	50.5	51	INS_BALPH	P01312 balaenopter
2	53	50.5	108	INS_PIG	P01315 sus scrofa
3	53	50.5	110	INS_CANFA	P01321 canis faml
4	53	50.5	110	INS_CERAE	P30407 cercopithec
5	53	50.5	110	INS_HUMAN	P01308 homo sapien
6	53	50.5	110	INS_MACFA	P30406 macaca fasc
7	53	50.5	110	INS_PANTR	P30410 pan troglod
8	53	50.5	110	INS_RABIT	P01311 ryctotlagus
9	50	47.6	51	INS_ACOCA	P01324 acomys cahi
10	50	47.6	51	INS_BALBO	P01314 balaenopter
11	50	47.6	51	INS_ELEMA	P01316 elephas max
12	50	47.6	108	INS1_MOUSE	P01325 mus musculus
13	50	47.6	108	INS_AOTTR	P06004 aotus trivi
14	50	47.6	110	INS1_RAT	P01322 rattus norv
15	50	47.6	110	INS2_MOUSE	P01326 mus musculus
16	50	47.6	110	INS2_RAT	P01323 rattus norv
17	50	47.6	110	INS_CRILO	P01313 cricetus
18	49	46.7	51	INS_CAMDR	P01320 camelus dro
19	49	46.7	86	INS_HORSE	P01310 equus cabal
20	49	46.7	105	INS_BOVIN	P01317 bos taurus
21	49	46.7	110	INS_PSAOB	Q62587 psammomys o
22	48	45.7	51	INS_HYSCR	P01328 hystrix cri
23	47	44.8	51	INS_CHIER	P01327 chinchilla
24	47	44.8	605	INS_TALEPL	Q92452 talaromyces
25	46	43.8	51	INS_ANSAN	P07454 anser anser
26	46	43.8	51	INS_CROAT	P01334 crotalus at
27	46	43.8	81	INS_ANAPL	P01333 anas platyr
28	46	43.8	515	INS1_BACST	Q8r185 bacillus st
29	46	43.8	555	INS1_PICAN	P21360 pichia angu
30	45	42.9	51	INS_ALLMI	P12703 alligator m
31	45	42.9	51	INS_CAPHI	P01319 capra hircu
32	45	42.9	51	INS_TRASC	P31887 trachemys s
33	45	42.9	103	INS_SELRF	P51463 selasphorus

ALIGNMENTS

RESULT 1	INS_BALPH	STANDARD;	PRT;	51 AA.
ID	INS_BALPH			
AC	P01312;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Insulin.			
GN	INS.			
OS	Balaenoptera physalus (Finback whale) (Common rorqual), and			P01318 ovis aries
OS	Physeter catodon (Sperm whale) (Physeter macrocephalus).			P01332 gallus gall
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			P06306 felis silve
OC	Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;			P12707 xenopus lae
OC	Balaenopteridae; Balaenoptera.			Q8y5r9 listeria m
OX	NCBI_TaxID=9770, 9755;			P30952 saccharomyc
RN	[1]			P81156 penicillium
RP	PARTIAL SEQUENCE.			P12708 zaocys dhum
RC	SPECIES=B.physalus;			P81423 acipenser g
RA	Hama H., Titani K., Sakaki S., Narita K.;			Q92a28 listeria in
RT	"The amino acid sequence in fin-whale insulin.";			P77947 streptomyc
RL	J. Biochem. 56:285-293(1964).			P18109 didelphis m
RN	[2]			
RP	SEQUENCE.			
RC	SPECIES=P.catodon;			
RA	Ishihara Y., Saito T., Ito Y., Fujino M.;			
RT	"Structure of sperm- and sei-whale insulins and their breakdown by			
RL	whale pepsin.";			
RN	Nature 181:1468-1469(1958).			
RP	SEQUENCE.			
RC	SPECIES=P.catodon;			
RA	Harris J.I., Sanger F., Naughton M.A.;			
RT	"Species differences in insulin.";			
RL	Arch. Biochem. Biophys. 65:427-438(1956).			
CC	-1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT			
CC	INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND			
CC	FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE			
CC	CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.			
CC	-1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO			
CC	DISULFIDE BONDS.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.			
DR	PIR; A91918; INWHP.			
DR	PIR; A93142; INWHP.			
DR	HSP; P01317; IAPH.			
DR	InterPro: IPR004825; Ins/IGF/relax.			
DR	PRINTS; PR00276; INSULINA.			
DR	PRINTS; PR00277; INSULINB.			
DR	SMART; SM00078; ILGF; 1.			
DR	PROSITE; PS00262; INSULIN; 1.			
KW	Insulin family; Hormone; Glucose metabolism.			
FT	CHAIN 1 30 INSULIN B CHAIN.			
FT	NON_CONS 30 31			
FT	CHAIN 31 51 INSULIN A CHAIN.			
FT	DISULFID 7 37 INTERCHAIN.			
FT	DISULFID 19 50 INTERCHAIN.			

FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5766 MW; 9007B514691A7CDD CRC64;
Query Match 50.5%; Score 53; DB 1; Length 51;
Best Local Similarity 55.0%; Pred. NO. 0.021;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 2 LVEQASTSQASLYQIYNFDN 21
DB 32 IVEQCCTSIQSLYQIYNFCN 51
RESULT 2
INS_PIG STANDARD; PRT; 108 AA.
AC P01315; Q9TFSJ5;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin precursor.
GN INS.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Han X.G., Tsch B.E.;
RT "Complete porcine preproinsulin cDNA sequence.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=68286485; PubMed=5657063;
RA Chance R.E., Ellis R.M., Bromer W.W.;
RT "Porcine proinsulin: characterization and amino acid sequence.";
RL Science 161:163-167(1968).
RN [3]
RP REVISION TO 59.
RA Chance R.E.;
RL Submitted (JUL-1970) to the PIR data bank.
RN [4]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;
RT "Insulin. The structure in the crystal and its reflection in chemistry and biology.";
RL Adv. Protein Chem. 26:279-402(1972).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RA Isaacs N.W., Agarwal R.C.;
RT "Experience with fast Fourier least squares in the refinement of the crystal structure of rhombohedral 2-zinc insulin at 1.5-A resolution.";
RL Acta Crystallogr. A 34:782-791(1978).
RN [6]
RX X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J., Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W., Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;
RT "The structure of 22n pig insulin crystals at 1.5-A resolution.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456(1988).
RN [7]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RA Balschmidt P., Hansen F.B., Dodson E., Dodson G., Korber F.;
RT "Structure of porcine insulin cocrystallized with clupeine 2.";
RL Acta Crystallogr. B 47:975-986(1991).
RN [8]
RX X-RAY CRYSTALLOGRAPHY.
RA Badger J., Harris M.R., Reynolds C.D., Evans A.C., Dodson E.J., Dodson G.G., North A.C.T.;
RT "Structure of the pig insulin dimer in the cubic crystal.";

RL Acta Crystallogr. B 47:127-136(1991).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RA Diaio J.-S., Wan Z.-L., Chang W.-R., Liang D.-C.;
RT "Structure of monomeric porcine DesBI-B2 despentapeptide (B26-B30) Insulin at 1.65-A resolution.";
RL Acta Crystallogr. D 53:507-512(1997).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -1- DATABASE: NAME-Protein Spotlight;
CC NOTE-Issue 9 of April 2001;
CC WWW="http://www.expasy.org/spotlight/articles/sptit009.html".
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF064555; AAC77920.1; ALT_INIT.
DR PIR; A01583; IPGC.
DR PDB; 3INS; 09-JAN-89.
DR PDB; 4INS; 31-JUL-94.
DR PDB; 6INS; 31-JAN-94.
DR PDB; 7INS; 31-JAN-94.
DR PDB; 9INS; 15-OCT-91.
DR PDB; 11ZA; 15-OCT-91.
DR PDB; 11ZB; 15-OCT-91.
DR PDB; 2TCI; 29-JAN-96.
DR PDB; 1MPJ; 29-JAN-96.
DR PDB; 3MTH; 29-JAN-96.
DR PDB; 1DEI; 16-JUN-97.
DR PDB; 1SDB; 01-APR-98.
DR PDB; 1WAV; 28-FEB-97.
DR PDB; 1ZEI; 16-FEB-99.
DR PDB; 1ZNI; 28-JAN-98.
DR PDB; 1ZNI; 28-JAN-98.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 85 C PEPTIDE.
FT CHAIN 88 108 INSULIN A CHAIN.
FT DISULFID 31 94 INTERCHAIN.
FT DISULFID 43 107 INTERCHAIN.
FT DISULFID 93 98
FT STRAND 26 46
FT STRAND 48 48
FT HELIX 89 94
FT HELIX 100 106
FT STRAND 107 107
SQ SEQUENCE 108 AA; 11671 MW; CB8491B429858EBE CRC64;
Query Match 50.5%; Score 53; DB 1; Length 108;
Best Local Similarity 55.0%; Pred. NO. 0.051;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 2 LVEQASTSQASLYQIYNFDN 21
DB 89 IVEQCCTSIQSLYQIYNFCN 108

RESULT 3
INS_CANFA INS_CANFA STANDARD; PRT; 110 AA.
AC P01321;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83109071; PubMed=6296142;
RA Kwok S.C.M., Chan S.J., Steiner D.F.;
RT Cloning and nucleotide sequence analysis of the dog insulin gene.
RT Coded amino acid sequence of canine preproinsulin predicts an
RT additional C-peptide fragment.*;
RL J. Biol. Chem. 258:2357-2363(1983).
RN [2]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.*;
RL Am. J. Med. 40:662-666(1966).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V00179; CAA23475.1; -
DR PIR; A01587; IPDG.
DR HSP; P01317; IAPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100 INTERCHAIN.
SQ SEQUENCE 110 AA; 12190 MW; A574791864A4FB98 CRC64;
Query Match 50.5%; Score 53; DB 1; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.052;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 2 LVEQSTQASLYQIYNEDN 21

DB 91 IVEQCCTSCSLYOLENYCN 110

RESULT 4
INS_CERAE INS_CERAE STANDARD; PRT; 110 AA.
AC P30407; P01309;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219953; PubMed=1560757;
RA Seino S., Bell G.I., Li W.;
RT "Sequences of primate insulin genes support the hypothesis of a
RT slower rate of molecular evolution in humans and apes than in
RT monkeys.*;
RL Mol. Biol. Evol. 9:193-203(1992).
RN [2]
RP SEQUENCE OF 57-87.
RX MEDLINE=72258016; PubMed=4626369;
RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
RT dog proinsulin C-peptides by a semi-micro Edman degradation
RT procedure.*;
RL J. Biol. Chem. 247:4866-4871(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X61092; CAA43405.1; -
DR PIR; A05232; A05232.
DR PIR; S22056; S22056.
DR PIR; B42179; B42179.
DR HSP; P01308; IAI0.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100 INTERCHAIN.
SQ SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;
Query Match 50.5%; Score 53; DB 1; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.052;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 2 LVQASTSQASLYQIYNFDN 21
DB. 91 IVEQCCTSCSYOLENYCN 110

RESULT 5
INS_HUMAN STANDARD; PRT; 110 AA.
AC P01308;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Insulin precursor.
GN INS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80120725; PubMed=6243748;
RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischler E.,
RA Goodman H.M.;
RT "Sequence of the human insulin gene.";
RL Nature 284:26-32(1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=80236313; PubMed=6248962;
RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
RT "Genetic variation in the human insulin gene.";
RL Science 209:612-615(1980).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80054779; PubMed=503234;
RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
RA Rutter W.J.;
RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin.";
RL Nature 282:525-527(1979).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=80147417; PubMed=6927840;
RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
RT "Nucleotide sequence of human preproinsulin complementary DNA.";
RL Science 208:57-59(1980).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=93364428; PubMed=8358440;
RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;
RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
kb segment of DNA spanning the insulin gene and associated VNTR.";
RL Nat. Genet. 4:305-310(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1-59 FROM N.A.
RC TISSUE=Blood;
RA Fajardy I.I., Weill J.J., Stuckens C.C., Danze P.M.P.;
RT "Description of a novel RFLP diallelic polymorphism (-127 BsgI C/G)
within the 5' region of insulin gene.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 25-54 AND 90-110.
RA Nicol D.S.H.W., Smith L.F.;
RT "Amino-acid sequence of human insulin.";
RL Nature 187:483-485(1960).
RN [9]
RP SEQUENCE OF 57-87.
RX MEDLINE=7116410; PubMed=5101771;
RA Oyer P.E., Cho S., Peterson J.D., Steiner D.F.;
RT "Studies on human proinsulin. Isolation and amino acid sequence of
the human pancreatic C-peptide.";
J. Biol. Chem. 246:1375-1386(1971).
RN [10]
RP SEQUENCE OF 57-87.
RX MEDLINE=71257722; PubMed=5560404;
RA Ko A., Smyth D.G., Markussen J., Sundby F.;
RT "The amino acid sequence of the C-peptide of human proinsulin.";
Eur. J. Biochem. 20:190-199(1971).
RN [11]
RP SYNTHESIS.
RX MEDLINE=75077277; PubMed=4443293;
RA Sieber P., Kamber B., Hartmann A., Joehli A., Riniker B., Rittel W.;
RT "Total synthesis of human insulin under directed formation of the
disulfide bonds.";
Helv. Chim. Acta 57:2617-2621(1974).
RN [12]
RP SYNTHESIS OF 57-87.
RX MEDLINE=75040007; PubMed=4803504;
RA Naithani V.K.;
RT "Studies on polypeptides, IV. The synthesis of C-peptide of human
proinsulin.";
Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).
RN [13]
RP SYNTHESIS OF 65-69 AND 70-73.
RX MEDLINE=73161263; PubMed=4698555;
RA Geiger R., Volk A.;
RT "Synthesis of peptides with the properties of human proinsulin C
peptides (hC peptide). 3. Synthesis of the sequences 14-17 and 9-13
of human proinsulin C peptides.";
Chem. Ber. 106:199-205(1973).
RN [14]
RP SYNTHESIS OF 84-87.
RX MEDLINE=73161261; PubMed=4698553;
RA Geiger R., Jaeger G., Keonig W., Treuth G.;
RT "Synthesis of peptides with the properties of human proinsulin C
peptides (hC peptide). I. Scheme for the synthesis and preparation of
the sequence 28-31 of human proinsulin C peptide.";
Chem. Ber. 106:188-192(1973).
RN [15]
RP VARIANT LOS ANGELES.
RX MEDLINE=84016053; PubMed=6312455;
RA Haneda M., Chan S.J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;
RT "Studies on mutant human insulin genes: Identification and sequence
analysis of a gene encoding [SerB24]insulin.";
Proc. Natl. Acad. Sci. U.S.A. 80:6366-6370(1983).
RN [16]
RP VARIANTS LOS ANGELES AND CHICAGO.
RX MEDLINE=84170233; PubMed=6424111;
RA Shoelson S., Fickova M., Haneda M., Nahum A., Musso G., Kaiser E.T.,
RA Rubenstein A.H., Tager H.;
RT "Identification of a mutant human insulin predicted to contain a
serine-for-phenylalanine substitution.";
Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).
RN [17]
RP VARIANT PROVIDENCE.
RX MEDLINE=87175640; PubMed=3470784;
RA Chan S.J., Seino S., Gruppaso P.A., Schwartz R., Steiner D.F.;
RT "A mutation in the B chain coding region is associated with impaired
proinsulin conversion in a family with hyperproinsulinemia.";
Proc. Natl. Acad. Sci. U.S.A. 84:2194-2197(1987).
RN [18]
RP VARIANT WAKAYAMA.
RX MEDLINE=87058122; PubMed=3537011;
RA Sakura H., Iwamoto Y., Sakamoto Y., Kuzuya T., Hirata H.;
RT "Structurally abnormal insulin in a diabetic patient. Characterization
of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";
J. Clin. Invest. 78:1666-1672(1986).
RN [19]
RP VARIANT HIS-89.
RX MEDLINE=90317021; PubMed=2196279;
RA Barbeti F., Raben N., Kadowaki T., Cama A., Accilli D., Gabbay K.H.,
RA Merenich J.A., Taylor S.I., Roth J.;
RT "Two unrelated patients with familial hyperproinsulinemia due to a

RT mutation substituting histidine for arginine at position 65 in the
RT proinsulin molecule: identification of the mutation by direct
RT sequencing of genomic deoxyribonucleic acid amplified by polymerase
RT chain reaction.";
RL J. Clin. Endocrinol. Metab. 71:164-169(1990).
RN [20]
RP VARIANT HIS-89.
RX MEDLINE-85261996; PubMed-4019786;
RA Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;
RT "Posttranslational cleavage of proinsulin is blocked by a point
RT mutation in familial hyperproinsulinemia.";
RL J. Clin. Invest. 76:378-380(1985).
RN [21]
RP VARIANT KYOTO.
RX MEDLINE-92291307; PubMed-1601997;
RA Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;
RT "A novel point mutation in the human insulin gene giving rise to
RT hyperproinsulinemia (proinsulin Kyoto).";
RL J. Clin. Invest. 89:1902-1907(1992).
RN [22]
RP STRUCTURE BY NMR.
RX MEDLINE-91104966; PubMed-2271664;
RA Hua Q.-X., Weiss M.A.;
RT "Toward the solution structure of human insulin: sequential 2D 1H NMR
RT assignment of a des-pentapeptide analogue and comparison with crystal
RT structure.";
RL Biochemistry 29:10545-10555(1990).
RN [23]
RP STRUCTURE BY NMR.
RX MEDLINE-91242467; PubMed-2036420;
RA Hua Q.-X., Weiss M.A.;
RT "Comparative 2D NMR studies of human insulin and des-pentapeptide
RT insulin: sequential resonance assignment and implications for protein
RT dynamics and receptor recognition.";
RL Biochemistry 30:5505-5515(1991).
RN [24]
RP STRUCTURE BY NMR.
RX MEDLINE-91265527; PubMed-1646635;
RA Hua Q.-X., Weiss M.A.;
RT "Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-
RT specific resonance assignments and effects of solvent composition.";
RL Biochim. Biophys. Acta 1078:101-110(1991).
RN [25]
RP STRUCTURE BY NMR.
RX MEDLINE-93059366; PubMed-1433291;
RA Joergensen A.M.M., Kristensen S.M., Led J.J., Balschmidt P.;
RT "Three-dimensional solution structure of an insulin dimer. A study of
RT the B9(Asp) mutant of human insulin using nuclear magnetic resonance,
RT distance geometry and restrained molecular dynamics.";
RL J. Mol. Biol. 227:1146-1163(1992).
RN [26]
RP STRUCTURE BY NMR OF VARIANT LOS-ANGELES.
RX MEDLINE-93133832; PubMed-8421693;
RA Hua Q.-X., Shoelson S.E., Inouye K., Weiss M.A.;
RT "Paradoxical structure and function in a mutant human insulin
RT associated with diabetes mellitus.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:582-586(1993).
RN [27]
RP STRUCTURE BY NMR.
RX MEDLINE-9738146; PubMed-9235985;
RA Chang X., Joergensen A.M., Bardrum P., Led J.J.;

Query Match 50.5%; Score 53; DB 1; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.052;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
:|||||:|:|
Db 91 IVEQCCTCSLYQLENYCN 110

RESULT 6
INS_MACFA

ID INS_MACFA STANDARD; PRT; 110 AA.
AC P30406; P01309;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-63080474; PubMed-6184262;
RA Wetekamp W., Groneberg J., Leineweber M., Wengenmayer F.;
RT "The nucleotide sequence of cDNA coding for preproinsulin from the
RT primate Macaca fascicularis.";
RL Gene 19:179-183(1982).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; J00336; AAA36849.1;
DR PIR: J00178; J00178.
DR HSP: P01308; IAI0.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PRO0276; INSULIN.
DR PRINTS: PRO0277; INSULIN.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA; 11991 MW; 83C6E33A80A420F9 CRC64;
Query Match 50.5%; Score 53; DB 1; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.052;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
OY 2 LVEQASTSQASLYQIYNFDN 21
:|||||:|:|
Db 91 IVEQCCTCSLYQLENYCN 110
RESULT 7
INS_PANTR
ID INS_PANTR STANDARD; PRT; 110 AA.
AC P30410;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.

```

GN OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219953; PubMed=1560757;
RA Seino S., Bell G.I., Li W.;
RT "Sequences of primate insulin genes support the hypothesis of a
RT slower rate of molecular evolution in humans and apes than in
RT monkeys.";
RL Mol. Biol. Evol. 9:193-203(1992).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61089; CA43403.1; -
DR PIR; S22058; S22058.
DR PIR; A42179; A42179.
DR HSSP; P01308; 1A10.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INSULIN A CHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100 INTERCHAIN.
SQ SEQUENCE 110 AA; 12025 MW; 41EB8DF79837CEF5 CRC64;

Query Match 50.5%; Score 53; DB 1; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.052;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYFNFDN 21
DB 91 IVEQCCTSGICSLYLENYCN 110

RESULT 8
INS_RABIT
ID INS_RABIT STANDARD; PRT; 110 AA.
AC P01311;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin precursor.
GN INS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-New Zealand white; TISSUE-Pancreas;
RX MEDLINE=94179230; PubMed=8132571;
RA Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R.,
RA Menon R.K., Zahm D.S.;
RT "Insulin gene expression and insulin synthesis in mammalian neuronal
RT cells.";
RL J. Biol. Chem. 269:8445-8454(1994).
RN [2]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.";
RL Am. J. Med. 40:662-666(1966).
RN [3]
RP SEQUENCE OF 56-110 FROM N.A.
RA Giddings S.J., Carnaghi L.R., Devaskar S.U.;
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03610; AA019033.1; -
DR EMBL; M61153; AA017540.1; -
DR PIR; A01581; INRB.
DR HSSP; P01308; 1TYM.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INSULIN A CHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100 INTERCHAIN.
FT CONFLICT 83 83 E -> Y (IN REF. 3).
SQ SEQUENCE 110 AA; 11838 MW; 82D2975B85D77FA8 CRC64;

Query Match 50.5%; Score 53; DB 1; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.052;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYFNFDN 21
DB 91 IVEQCCTSGICSLYLENYCN 110

RESULT 9
INS_ACOCA
ID INS_ACOCA STANDARD; PRT; 51 AA.
AC P01324;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Insulin.

```

GN INS.
OS Acomys cahirinus (Egyptian spiny mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Acomys.
OX NCBI_TaxID=10060;
[1]
RN COMPOSITION.
RP MEDLINE=72189454; PubMed=5028210;
RA Buenzli H.F., Humbel R.E.;
RT "Isolation and partial structural analysis of insulin from mouse (Mus musculus) and spiny mouse (Acomys cahirinus).";
RL Hoppe-Seyler's Z. Physiol. Chem. 353:444-450(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01591; INMSP.
DR HSP: P01308; IYTM.
DR InterPro: IPR004825; Ins/IGF/relax.
DR PRINTS: PR00276; INSULIN.
DR SMART: SM00078; ILGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5768 MW; 992BD8629047D3D CRC64;
Query Match 47.6%; Score 50; DB 1; Length 51;
Best Local Similarity 50.0%; Pred. No. 0.066;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
[1]
OY 2 LVEQASTSQASLYQIYNFDN 21
Db 32 IVDQCTSCSYQLENYCN 51
[1]
RESULT 10
INS_BALBO STANDARD; PRT; 51 AA.
AC P01314;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE Insulin.
GN INS.
OS Balaenoptera borealis (Sei whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9768;
[1]
RN COMPOSITION.
RP Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin.";
RL Nature 181:1468-1469(1958).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

DR PIR: A01582; INWHLS.
DR HSP: P01317; IAPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR PRINTS: PR00276; INSULIN.
DR SMART: SM00078; ILGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;
Query Match 47.6%; Score 50; DB 1; Length 51;
Best Local Similarity 50.0%; Pred. No. 0.066;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
[1]
OY 2 LVEQASTSQASLYQIYNFDN 21
Db 32 IVEQCASTCSLYQLENYCN 51
[1]
RESULT 11
INS_ELEMA STANDARD; PRT; 51 AA.
AC P01316;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE Insulin.
GN INS.
OS Elephas maximus (Indian elephant).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Proboscidea; Elephantiidae; Elephas.
OX NCBI_TaxID=9783;
[1]
RN SEQUENCE.
RP MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.";
RL Am. J. Med. 40:662-666(1966).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THE SPECIES OF ELEPHANT IS NOT GIVEN, BUT IT IS MOST PROBABLY THE INDIAN ELEPHANT (ELEPHAS MAXIMUS).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01584; INEL.
DR HSP: P01308; IAI0.
DR InterPro: IPR004825; Ins/IGF/relax.
DR PRINTS: PR00276; INSULIN.
DR SMART: SM00078; ILGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5752 MW; 9007B50CDB457D6D CRC64;
Query Match 47.6%; Score 50; DB 1; Length 51;
Best Local Similarity 50.0%; Pred. No. 0.066;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

CC CC 2 LVEQASTQASLYQIYNFDN 21
 DB 32 IVEQCTGVCGLYQLENTCN 51
 RESULT 12
 ID INSL_MOUSE STANDARD; PRT; 108 AA.
 AC P01325; Q9D907;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Insulin 1 precursor.
 GN INSL OR INS-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87169768; PubMed=3104603;
 RA Wentworth B.M., Schaefer I.M., Villa-Komaroff L., Chirgwin J.M.;
 RT "Characterization of the two nonallelic genes encoding mouse
 preproinsulin.";
 RL J. Mol. Evol. 23:305-312(1986).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90372989; PubMed=2397023;
 RA Sawa T., Ohgaku S., Morioka H., Yano S.;
 RT "Molecular cloning and DNA sequence analysis of preproinsulin genes
 in the NON mouse, an animal model of human non-obese, non-insulin-
 dependent diabetes mellitus.";
 RL J. Mol. Endocrinol. 5:61-67(1990).
 [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=C57BL/6J; TISSUE=Pancreas;
 MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 [4]
 RP SEQUENCE OF 25-54 AND 88-108.
 RX MEDLINE=72189455; PubMed=5063718;
 RA Buenzli H.F., Glatthaar B., Kunz P., Muelhaupt E., Humbel R.E.;
 RT "Amino acid sequence of the two insulins from mouse (Mus musculus).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 353:451-458(1972).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; X04725; CAA28434.1; -
 DR EMBL; AK007482; BAB25058.1; -
 DR PIR; A01592; INMS1.
 DR PIR; B26342; B26342.
 DR PIR; A48172; A48172.
 DR HSSP; P01308; 1A7F.
 DR MGD; MGI:96572; Ins1.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00276; INSULIN.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism; Signal; Multigene family.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN 1 B CHAIN.
 FT PROPEP 57 85 INSULIN 1 C PEPTIDE.
 FT CHAIN 88 108 INSULIN 1 A CHAIN.
 FT DISULFID 31 94 INTERCHAIN.
 FT DISULFID 43 107 INTERCHAIN.
 FT DISULFID 93 98
 FT CONFLICT 37 37 E -> K (IN REF. 3).
 SQ SEQUENCE 108 AA; 12160 MW; F63D9B7B896E0F88 CRC64;
 Query Match 47.6%; Score 50; DB 1; Length 108;
 Best Local Similarity 50.0%; Pred. No. 0.16; Mismatches 6; Indels 0; Gaps 0;
 Matches 10; Conservative 4;
 OY 2 LVEQASTQASLYQIYNFDN 21
 DB 89 IVDQCTGVCGLYQLENTCN 108
 RESULT 13
 ID INS_AOTTR STANDARD; PRT; 108 AA.
 AC P10604;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Insulin precursor.
 GN INS.
 OS Aotus trivirgatus (Night monkey) (Douroucoul), and
 OS Saimiri sciureus (Common squirrel monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
 OX NCBI_TaxID=9505, 9521;
 [1]
 RN SEQUENCE FROM N.A.
 RP SPECIES=A.trivirgatus;
 RX MEDLINE=88041119; PubMed=3118367;
 RA Seino S., Steiner D.F., Bell G.I.;
 RT "Sequence of a New World primate insulin having low biological
 potency and immunoreactivity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7423-7427(1987).
 [2]
 RN SEQUENCE OF 25-54 AND 88-108.
 RP SPECIES=S.sciureus;
 RX MEDLINE=91088593; PubMed=2263627;
 RA Yu J.-H., Eng J., Yalow R.S.;
 RT "Isolation and amino acid sequences of squirrel monkey (Saimiri
 sciurea) insulin and glucagon.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9766-9768(1990).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT

```
CC CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC CC DISULFIDE BONDS.
CC CC -1- SUBCELLULAR LOCATION: Secreted.
CC CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; V01243; CAA24560.1; -
CC CC EMBL; J00748; AAA1443.1; -
CC CC EMBL; M25585; AAA1440.1; -
CC CC EMBL; M25583; AAA1440.1; JOINED.
CC CC PIR; A01590; IPRT2.
CC CC HSSP; P01317; IAPH.
CC CC InterPro; IPR004825; Ins/IGF/relax.
CC CC Pfam; PF00049; Insulin; 1.
CC CC PRINTS; PR00276; INSULIN.
CC CC PRINTS; PR00277; INSULIN.
CC CC SMART; SM00078; ILGF; 1.
CC CC PROSITE; PS00262; INSULIN; 1.
CC CC Insulin family; Hormone; Glucose metabolism; Signal; Multigene family.
CC CC SIGNAL 1 24
CC CC CHAIN 25 54 INSULIN 2 B CHAIN.
CC CC PROPEP 57 87 INSULIN 2 C PEPTIDE.
CC CC CHAIN 90 110 INSULIN 2 A CHAIN.
CC CC DISULFID 31 96 INTERCHAIN.
CC CC DISULFID 43 109 INTERCHAIN.
CC CC DISULFID 95 100 INTERCHAIN.
CC CC SEQUENCE 110 AA; 12339 MW; 3A626DA98C86F3CA CRC64;

Query Match 47.6%; Score 50; DB 1; Length 110;
Best Local Similarity 50.0%; Pred. No. 0.17;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Oy 2 LVEQASTSQASLYQIYNFDN 21
Db 91 IVDQCCTCSLYQLENYCN 110

RESULT 17
INS_CRILQ
ID INS_CRILQ STANDARD; PRT; 110 AA.
AC P01313;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus
OC Cricetulus
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84133036; PubMed=6365663;
RA Bell G.I., Sanchez-Pescador R.;
FT "Sequence of a cDNA encoding Syrian hamster preproinsulin."
RL Diabetes 33:297-300(1984).
RN [2]
RP SEQUENCE OF 25-54 AND 90-110.
RA Neelon F.A., Delcher H.K., Steinman H., Lebovitz H.E.;
FT "Structure of hamster insulin: comparison with a tumor insulin."
RL Fed. Proc. 32:300-300(1973).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC
```

```
CC CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC CC DISULFIDE BONDS.
CC CC -1- SUBCELLULAR LOCATION: Secreted.
CC CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; M26328; AAA37089.1; -
CC CC PIR; A91456; INHY.
CC CC HSSP; P01308; ITHM.
CC CC InterPro; IPR004825; Ins/IGF/relax.
CC CC Pfam; PF00049; Insulin; 1.
CC CC PRINTS; PR00276; INSULIN.
CC CC PRINTS; PR00277; INSULIN.
CC CC SMART; SM00078; ILGF; 1.
CC CC PROSITE; PS00262; INSULIN; 1.
CC CC Insulin family; Hormone; Glucose metabolism; Signal.
CC CC SIGNAL 1 24
CC CC CHAIN 25 54 INSULIN B CHAIN.
CC CC PROPEP 57 87 C PEPTIDE.
CC CC CHAIN 90 110 INSULIN A CHAIN.
CC CC DISULFID 31 96 INTERCHAIN.
CC CC DISULFID 43 109 INTERCHAIN.
CC CC DISULFID 95 100 INTERCHAIN.
CC CC SEQUENCE 110 AA; 12268 MW; 219E92B85A535C8C CRC64;

Query Match 47.6%; Score 50; DB 1; Length 110;
Best Local Similarity 50.0%; Pred. No. 0.17;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Oy 2 LVEQASTSQASLYQIYNFDN 21
Db 91 IVDQCCTCSLYQLENYCN 110

RESULT 18
INS_CAMDR
ID INS_CAMDR STANDARD; PRT; 51 AA.
AC P01320;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838;
RN [1]
RP SEQUENCE.
RA Danho W.O.;
FT "The isolation and characterization of insulin of camel (Camelus
FT dromedarius)."
RL J. Fac. Med. Baghdad 14:16-28(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC CC DISULFIDE BONDS.
CC CC -1- SUBCELLULAR LOCATION: Secreted.
CC CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC CC PIR; A92782; INCM.
CC CC HSSP; P01317; 2INS.
CC CC InterPro; IPR004825; Ins/IGF/relax.
CC CC PRINTS; PR00276; INSULIN.
CC CC PRINTS; PR00277; INSULIN.
```


DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON CONS 30 31
FT CHAIN 31 31 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5693 MW; 901E88BA085A7DDD CRC64;

Query Match 46.7%; Score 49; DB 1; Length 51;
Best Local Similarity 50.0%; Pred. No. 0.098;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 LVEQASTSOASLYQIYNFDN 21
:|||||:|:|
Db 32 IVEQCCASVCSLYQENYCN 51

RESULT 19
INS_HORSE
ID INS_HORSE STANDARD; PRT; 86 AA.
AC P01310;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Insulin precursor.
GN INS.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE OF 1-30 AND 66-86.
RA Harris J.I., Sanger F., Naughton M.A.;
RT "Species differences in insulin."
RL Arch. Biochem. Biophys. 65:427-438(1956).
RN [2]
RP SEQUENCE OF 33-63.
RX MEDLINE-73061498; PubMed-4640931;
RA Tager H.S., Steiner D.F.;
RT "Primary structures of the proinsulin connecting peptides of the rat and the horse."
RL J. Biol. Chem. 247:7936-7940(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -1- CAUTION: X'S AT POSITIONS 31-32 AND 64-65 REPRESENT PAIRED BASIC RESIDUES ASSUMED BY HOMOLGY TO BE PRESENT IN THE PRECURSOR MOLECULE.
DR PIR; A01580; IPHO.
DR HSP; P01317; IAPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT PROPEP 33 63 C PEPTIDE.
FT CHAIN 66 86 INSULIN A CHAIN.
FT DISULFID 7 72 INTERCHAIN.
FT DISULFID 19 85 INTERCHAIN.
FT DISULFID 71 76
SQ SEQUENCE 86 AA; 9142 MW; A3E1E822711BDB46 CRC64;

Query Match 46.7%; Score 49; DB 1; Length 86;
Best Local Similarity 50.0%; Pred. No. 0.18;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 LVEQASTSOASLYQIYNFDN 21
:|||||:|:|
Db 67 IVEQCTGICSLYQENYCN 86

RESULT 20
INS_BOVIN
ID INS_BOVIN STANDARD; PRT; 105 AA.
AC P01317;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin precursor.
GN INS.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-88288209; PubMed-2456452;
RA D'Agostino J., Younes M.A., White J.W., Besch P.K., Field J.B., Frazier M.L.;
RT "Cloning and nucleotide sequence analysis of complementary deoxyribonucleic acid for bovine preproinsulin."
RL Mol. Endocrinol. 1:327-331(1987).
RN [2]
RP SEQUENCE OF 25-105.
RX MEDLINE-71166442; PubMed-4928892;
RA Nolan C., Margoliash E., Peterson J.D., Steiner D.F.;
RT "The structure of bovine proinsulin."
RL J. Biol. Chem. 246:2780-2795(1971).
RN [3]
RP SEQUENCE OF 25-54.
RA Sanger F., Tuppy H.;
RT "The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates."
RL Biochem. J. 49:481-490(1951).
RN [4]
RP SEQUENCE OF 57-82.
RX MEDLINE-71116409; PubMed-5545080;
RA Steiner D.F., Cho S., Oyer P.E., Terris S., Peterson J.D., Rubenstein A.H.;
RT "Isolation and characterization of proinsulin C-peptide from bovine pancreas."
RL J. Biol. Chem. 246:1365-1374(1971).
RN [5]
RP SEQUENCE OF 57-82.
RX MEDLINE-71257721; PubMed-5105368;
RA Salokangas A., Smyth D.G., Markussen J., Sundby F.;
RT "Bovine proinsulin: amino acid sequence of the C-peptide isolated from pancreas."
RL Eur. J. Biochem. 20:183-189(1971).
RN [6]
RP SEQUENCE OF 85-105.
RA Sanger F., Thompson E.O.P.;
RT "The amino-acid sequence in the glycyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates."
RL Biochem. J. 53:366-374(1953).
RN [7]
RP AMIDES, SEQUENCE OF 25-54 AND 85-105, AND DISULFIDE BONDS.
RA Ryle A.P., Sanger F., Smith L.F., Kitai R.;
RT "The disulphide bonds of insulin."
RL Biochem. J. 60:541-556(1955).
RN [8]
RP X-RAY CRYSTALLOGRAPHY.
RA Smith G.D., Duax W.L., Dodson E.J., Dodson G.G., de Graaf R.A.G., Reynolds C.D.;

```
RT "The structure of des-Phe b1 bovine insulin.";
RL Acta Crystallogr. B 38:3028-3032(1982).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.3 ANGSTROMS).
RX MEDLINE=97285914; PubMed=9141131;
RA Brande J., Dodson G.G., Edwards D.J., Holden P.H., Whittingham J.L.;
RT "A model of insulin fibrils derived from the X-ray crystal structure
of a monomeric insulin (despentapeptide insulin).";
RL Proteins 27:507-516(1997).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -1- DATABASE: NAMP-Protein Spotlight;
CC NOTE=Issue 9 of April 2001;
CC WWW="http://www.expasy.org/spotlight/articles/splt009.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M54979; AAA30722.1;
DR PIR; A01585; IPBO.
DR PIR; A40909; A40909.
DR PDB; 2INS; 31-MAY-84.
DR PDB; 1APH; 31-OCT-93.
DR PDB; 1BPH; 31-OCT-93.
DR PDB; 1CPH; 31-OCT-93.
DR PDB; 1DPH; 31-OCT-93.
DR PDB; 1PID; 07-DEC-96.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
DR Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 1 24 INSULIN B CHAIN.
FT PROPEP 25 54 C PEPTIDE.
FT CHAIN 57 82 INSULIN A CHAIN.
FT CHAIN 85 105 INTERCHAIN.
FT DISULFID 31 91 INTERCHAIN.
FT DISULFID 43 104 INTERCHAIN.
FT TURN 90 95
FT TURN 32 32
FT HELIX 33 46
FT STRAND 48 48
FT HELIX 86 90
FT TURN 91 94
FT HELIX 97 101
FT TURN 102 103
FT STRAND 104 104
SQ SEQUENCE 105 AA; 11393 MW; 75307CF78E61C06A CRC64;

Query Match 46.7%; Score 49; DB 1; Length 105;
Best Local Similarity 50.0%; Pred. No. 0.23;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||:
Db 86 IVEQCCASVCISLYOLENYCN 105

RESULT 21
INS_PSAOB
```

```
INS_PSAOB STANDARD; PRT; 110 AA.
AC QG2587;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Insulin precursor.
GN INS.
OS Psammomys obesus.
OC Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Psammomys.
OX NCBI_TaxID=48139;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97309250; PubMed=9166665;
RA Kaiser N., Ballyes E.M., Schneider B.S., Cerasi E., Steiner D.F.,
RA Hutton J.C., Gross D.J.;
RT "Characterization of the unusual insulin of Psammomys obesus, a
RT rodent with nutrition-induced NIDDM-like syndrome.";
RL Diabetes 46:953-957(1997).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X98241; CAA66897.1;
DR HSSP; P01308; 1A10.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
DR Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).
FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).
FT DISULFID 95 100 BY SIMILARITY.
SQ SEQUENCE 110 AA; 12324 MW; A006738E20579CB0 CRC64;

Query Match 46.7%; Score 49; DB 1; Length 110;
Best Local Similarity 50.0%; Pred. No. 0.24;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||:
Db 91 IVEQCCGICSLYOLENYCN 110

RESULT 22
INS_HYSCR STANDARD; PRT; 51 AA.
AC P01328;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin.
```

GN INS.
OS Hystrix cristata (Crested porcupine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Hystricidae; Hystrix.
OX NCBI_TaxID=10137;
RN [1]
RP SEQUENCE.
RX MEDLINE=80254705; PubMed=6995860;
RA Horuk R., Blundell T.L., Lazarus N.R., Neville R.W.J., Stone D.,
RA Wollmer A.;
RT "A monomeric insulin from the porcupine (Hystrix cristata), an Old
World hystricomorph.";
RL Nature 286:822-824(1980).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR; A01594; INPQ.
DR HSSP; P01308; ILPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN (BY SIMILARITY).
FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).
FT DISULFID 36 41 BY SIMILARITY.
FT SEQUENCE 51 AA; 5705 MW; 6C36087CC689161B CRC64;
Query Match 45.7%; Score 48; DB 1; Length 51;
Best Local Similarity 45.0%; Pred. No. 0.14; Mismatches 4; Indels 7; Gaps 0;
Matches 9; Conservative 4;
QY 2 LVEQASTSQASLYQIYNFDN 21
DB 32 IVDQCTGVCGLYQLENYCN 51
RESULT 23
INS_CHIBR STANDARD; PRT; 51 AA.
AC P01327;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Chinchilla brevicaudata (Chinchilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Chinchillidae;
OC Chinchilla.
OX NCBI_TaxID=10152;
RN [1]
RP SEQUENCE.
RX MEDLINE=76022416; PubMed=1175610;
RA Wood S.P., Blundell T.L., Wollmer A., Lazarus N.R., Neville R.W.J.;
RT "The relation of conformation and association of insulin to receptor
binding; x-ray and circular-dichroism studies on bovine and
hystricomorph insulins.";
RL Eur. J. Biochem. 55:531-542(1975).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO

CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR; A01593; INCB.
DR HSSP; P01308; IATF.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN (BY SIMILARITY).
FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).
FT DISULFID 36 41 BY SIMILARITY.
FT SEQUENCE 51 AA; 5741 MW; 8F7EC904691A78A0 CRC64;
Query Match 44.8%; Score 47; DB 1; Length 51;
Best Local Similarity 45.0%; Pred. No. 0.21; Mismatches 5; Indels 6; Gaps 0;
Matches 9; Conservative 5;
QY 2 LVEQASTSQASLYQIYNFDN 21
DB 32 IVDQCTGVCGLYQLENYCN 51
RESULT 24
GOX_TALFL STANDARD; PRT; 605 AA.
ID GOX_TALFL
AC Q92452;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glucose oxidase precursor (EC 1.1.3.4) (Glucose oxyhydrase) (GOD)
DE (Beta-D-glucose:oxygen 1-oxido-reductase).
GN GOX.
OS Talaromyces flavus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Talaromyces.
OX NCBI_TaxID=5095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 32908;
RX MEDLINE=98041929; PubMed=9371889;
RA Murray P.R., Llewellyn D.J., Peacock W.J., Dennis E.S.;
RT "Isolation of the glucose oxidase gene from Talaromyces flavus and
characterisation of its role in the biocontrol of Verticillium
dahliae.";
RL Curr. Genet. 32:367-375(1997).
CC -1- CATALYTIC ACTIVITY: Beta-D-glucose + O(2) -> D-glucono-1,5-lactone
+ H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; U56240; AAB09442.1;
DR HSSP; P81156; IGPE.
DR InterPro; IPR000172; GMC_oxred.
DR Pfam; PF00732; GMC_oxred; 1.
DR PROSITE; PS00623; GMC_OXRED.1; 1.
DR PROSITE; PS00624; GMC_OXRED.2; 1.
KW Oxidoreductase; Flavoprotein; FAD; Glycoprotein; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 605
GLUCOSE OXIDASE.

```

FT NP_BIND 44 73 FAD (ADP PART) (PROBABLE).
FT ACT_SITE 543 543 POTENTIAL.
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 605 AA; 65769 MW; 746F5F7FD1558D3 CRC64;

Query Match 44.8%; Score 47; DB 1; Length 605;
Best Local Similarity 57.9%; Pred. No. 4;
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 NLVEQASTSQASLYQIYNF 19
DB 515 NLVENATLSQWSYVLQNF 533

RESULT 25
INS_ANSAN
ID INS_ANSAN STANDARD; PRT; 51 AA.
AC P07454; Q10995;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Insulin.
GN INS.
OS Anser anser anser (Western graylag goose), and
OS Cairina moschata (Muscovy duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anser.
OX NCBI_TaxID=8844, 8855;
RN [1]
RP SEQUENCE.
RC SPECIES=A. anser;
RA Xu Y., Lin N., Zhang Y., Zhang Y.;
RT "Isolation and sequence determination of goose insulin.";
RL Xueue tongbao 28:966-968(1983).
RN [2]
RP SEQUENCE.
RC SPECIES=C. moschata;
RX MEDLINE=96321314; PubMed=8759296;
RA Chevallier B., Anglade P., Derouet M., Molle D., Simon J.;
RT "Isolation and characterization of Muscovy (Cairina moschata) duck
insulin.";
RL Comp. Biochem. Physiol. 114B:19-26(1996).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR; JC0007; INGS.
DR HSP; P01308; IHIS.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5716 MW; 976EFAED8C68386D CRC64;

Query Match 43.8%; Score 46; DB 1; Length 51;

```

```

Best Local Similarity 45.0%; Pred. No. 0.31;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
DB 32 IVEQCCENPCSLYQLENYCN 51

```

Search completed: December 23, 2002, 07:14:41
Job time : 7.17647 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:17 ; Search time 20.5882 Seconds
(without alignments)
210.168 Million cell updates/sec

Title: US-09-574-443-7_COPY_1_21
Perfect score: 105
Sequence: 1 NLVEQASTSQASLYQIYNFDN 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_21:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_undefined:*
 - 15: sp_virus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	50.5	110	11 Q91X13	Q91X13 spermophilu
2	50	47.6	39	11 Q62542	Q62542 mus spretus
3	50	47.6	41	11 Q62543	Q62543 mus spretus
4	47	44.8	256	17 Q8TWW4	Q8TWW4 methanosarc
5	47	44.8	1142	5 Q95XD2	Q95XD2 caenorhabdi
6	46	43.8	515	2 Q8RL85	Q8RL85 bacillus st
7	45.5	43.3	95	16 Q928D3	Q928D3 listeria in
8	44	41.9	110	6 Q8WNW6	Q8WNW6 felis silve
9	44	41.9	301	8 Q33762	Q33762 allomyces m
10	44	41.9	418	13 Q90XX5	Q90XX5 oncorhynch
11	44	41.9	418	13 Q90XX4	Q90XX4 oncorhynch
12	44	41.9	512	16 Q8Y5R9	Q8Y5R9 listeria mo
13	44	41.9	553	17 Q8TH40	Q8TH40 methanosarc
14	44	41.9	561	5 Q960S3	Q960S3 drosophila
15	44	41.9	605	3 Q9URJ8	Q9URJ8 penicillium
16	44	41.9	761	5 Q9VB04	Q9VB04 drosophila

17	43.5	41.4	545	5 Q9NEK9	Q9NEK9 caenorhabdi
18	43	41.0	151	12 O10260	O10260 human papil
19	43	41.0	271	5 P90946	P90946 caenorhabdi
20	43	41.0	512	16 Q92A28	Q92A28 listeria in
21	43	41.0	540	2 Q9ALX7	Q9ALX7 streptomyce
22	43	41.0	540	16 Q9RK09	Q9RK09 streptomyce
23	43	41.0	714	9 Q9XJK1	Q9XJK1 bacterioph
24	43	41.0	714	16 Q9KXC6	Q9KXC6 escherichia
25	43	41.0	732	5 Q24557	Q24557 drosophila
26	43	41.0	732	5 Q960X5	Q960X5 drosophila
27	43	41.0	732	5 Q24558	Q24558 drosophila
28	43	41.0	1144	2 Q87582	Q87582 mycoplasma
29	42.5	40.5	1139	5 Q21049	Q21049 caenorhabdi
30	42.5	40.5	1345	5 Q96236	Q96236 plasmodium
31	42	40.0	148	12 O11400	O11400 rhesus papi
32	42	40.0	166	10 Q39149	Q39149 arabidopsis
33	42	40.0	168	11 Q35065	Q35065 mus musculu
34	42	40.0	250	2 Q93UC8	Q93UC8 carsonella
35	42	40.0	269	10 Q39148	Q39148 arabidopsis
36	42	40.0	273	2 Q99QJ1	Q99QJ1 vibrio chol
37	42	40.0	274	5 Q9U387	Q9U387 caenorhabdi
38	42	40.0	528	2 Q8VM95	Q8VM95 alcaligenes
39	42	40.0	550	4 Q9NTN2	Q9NTN2 homo sapien
40	42	40.0	604	4 Q9NQZ7	Q9NQZ7 homo sapien
41	42	40.0	606	11 Q9ET10	Q9ET10 mus musculu
42	42	40.0	628	3 Q94732	Q94732 schizosacch
43	42	40.0	716	16 Q8UBR8	Q8UBR8 agrobacteri
44	41.5	39.5	195	16 Q8RH29	Q8RH29 fusobacteri
45	41.5	39.5	215	16 Q9S4J1	Q9S4J1 streptococc

ALIGNMENTS

RESULT 1

Q91X13 PRELIMINARY; PRT; 110 AA.

ID Q91X13; PRELIMINARY; PRT; 110 AA.

AC Q91X13; PRELIMINARY; PRT; 110 AA.

DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)

DE Insulin.

OS *Spermophilus tridecemlineatus* (Thirteen-lined ground squirrel).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sclurognathi; Scluridae; Sclurinae;

OC *Spermophilus*.

OX NCBI_TaxID=43179;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PANCREAS;

RA Tredrea M.M., Buck M.J., Guhaniyogi J., Squire T.L., Andrews M.T.;

RT "Regulation of PDK4 expression in a hibernating mammal.";

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

DR EMBL: AY038604; AAK72558.1;

DR InterPro: IPR004825; Ins/IGF/relax.

DR Pfam: PF00049; Insulin; 1.

DR PROSITE: PS00262; INSULIN; UNKNOWN_1.

SQ SEQUENCE 110 AA; 12004 MW; 4511768D6622BEE5 CRC64;

Query Match 50.5%; Score 53; DB 11; Length 110;
Best Local Similarity 55.0%; Pred. No. 0.46; Indels 0; Gaps 0;
Matches 11; Conservative 3; Mismatches 6;

QY 2 LVEQASTSQASLYQIYNFDN 21
Db 91 IVEQCTSIQSXYQLENYCN 110

RESULT 2

Q62542 PRELIMINARY; PRT; 39 AA.

ID Q62542

```

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: U05730; AAB60474.1; -
DR HSSP: P01308; 1A7F.
DR MGD: MGI:96573; Ins2.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULINA.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Hormone; Glucose metabolism; Multigene family.
FT NON_TER 1
FT CHAIN 21 41 A CHAIN.
SQ SEQUENCE 41 AA; 4361 MW; 55CDB871FF720672 CRC64;

Query Match 47.6%; Score 50; DB 11; Length 41;
Best Local Similarity 50.0%; Pred. No. 0.5;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYOIYNFDN 21
Db 22 IVQDCTISCLYQLENYCN 41
: : : : :
: : : : :

RESULT 4
Q8TTW4 PRELIMINARY; PRT; 256 AA.
ID Q8TTW4 AC
DT Q8TTW4; 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane protein.
GN MA0311.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OC NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZA / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Firrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RA "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL: AB010690; AAM03764.1; -
DR Complete proteome.
KW SEQUENCE 256 AA; 27332 MW; 5237CE0D205309D2 CRC64;

Query Match 44.8%; Score 47; DB 17; Length 256;
Best Local Similarity 41.2%; Pred. No. 11;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 5 QASTSQASLYOIYNFDN 21
Db 106 EMKTSYVSVYPTNYEN 122
: : : : :
: : : : :

RESULT 5
Q95XD2 PRELIMINARY; PRT; 1142 AA.
ID Q95XD2
```

Qy	1	NLVEQASTSQASLIQIYN 18	
	1	:::::1 1:1 :	
Db	417	NEIQEAATGSGSVYALYN 434	
RESULT 7			
Q928D3			
ID	Q928D3	PRELIMINARY;	PRT; 95 AA.
AC	Q928D3;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Hypothetical protein lin2603.		
GN	Lin2603.		
OS	Listeria innocua.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacilli		
OC	Listeriaceae; Listeria.		
OX	NCBI_TaxID=1642;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CLIP 11262 / SEROVAR 6A;		
RX	PubMed=11679669;		
RA	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend		
RA	Baquero F., Berche P., Bloeker H., Brandt P., Chakrabo		
RA	Charbit A., Chetoui F., Couve E., de Daruvar A., Dehou		
RA	Domann E., Dominguez-Bernal G., Duchaud E., Durant L., D		
RA	Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido		
RA	Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J.,		
RA	Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kur		
RA	Madueno E., Maitournan A., Mata Vicente J., Ng E., Nedja		
RA	Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C.,		
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,		
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;		
RT	*Comparative genomics of Listeria species.*;		
RL	Science 294:949-952(2001).		
DR	ENBL: AL596173; CAC97830.1;		
DR	Listlist; LIN02603; -		
DR	Hypothetical protein; Complete proteome.		
SK	SEQUENCE 95 AA; 11173 MW; 569A72BAAC4274CF CRC64;		
Qy	Query Match	43.3%;	Score 45.5; DB 16; Length
	Best Local Similarity	45.5%;	Pred. No. 6.8;
	Matches 10; Conservative	4;	Mismatches 7; Indels
Qy	1	NLVEQASTSQASLIQ-IYNFDN 21	
	1	:::::1 1:1 :	
Db	15	HILEQAENDQFGIOGTIYKFDN 36	
RESULT 8			
Q8WNW6			
ID	Q8WNW6	PRELIMINARY;	PRT; 110 AA.
AC	Q8WNW6;		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Preproinsulin.		
DE	Felis silvestris catus (Cat).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eute		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis		
OC	NCBI_TaxID=9685;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=PANCREAS;		
RA	Okamoto S., Morimatsu M.;		
RA	"Cat insulin.";		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases		
DR	ENBL: AB043535; BAB84110.1; -		
DR	Interpro: IPR004825; Ins/IGF/relax.		
DR	Pfam: PF00049; Insulin; 1.		
DR	PRINTS; PR00276; INSULINA.		
DR	PRINTS; PR00277; INSULINB.		

```
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; UNKNOWN_1.
SQ SEQUENCE 110 AA; 12069 MW; 95FB6E170C7BECA4 CRC64;

Query Match 41.9%; Score 44; DB 6; Length 110;
Best Local Similarity 45.0%; Pred. No. 14;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
Db 91 IVEQCCASVCSLYQLSHYCN 110

RESULT 9
O33762 PRELIMINARY; PRT; 301 AA.
AC Q33762;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE ORf301 protein.
GN ORF301.
OS Allomyces macrogynus.
OG Mitochondrion.
OC Eukaryota; Fungi; Chytridiomycota; Blastocladales; Blastocladiaceae;
OC Allomyces.
OX NCBI_TaxID=28583;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96226032; PubMed=8636971;
RA Paquin B., Lang B.F.;
RT "The mitochondrial DNA of Allomyces macrogynus: the complete genomic
sequence from an ancestral fungus.";
RL J. Mol. Biol. 255:688-701(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Paquin B., Laforest M.J., Lang B.F.;
RT "80 conserved, palindromic sequence elements in the mitochondrial DNA
of Allomyces sp.: Evidence for mobility.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U01288; AAC49244.1;
DR InterPro; IPR003647; Intron_nuc_1.
DR InterPro; IPR003611; Intron_nuc_2.
DR InterPro; IPR000305; UvrC_N.
DR Pfam; PF01541; Excl_endo_N; 1.
DR SMART; SM00465; GYVC; 1.
DR SMART; SM00497; IENR1; 1.
DR SMART; SM00496; IENR2; 3.
KW Mitochondrion.
SQ SEQUENCE 301 AA; 34210 MW; F696732239DFCBAA CRC64;

Query Match 41.9%; Score 44; DB 8; Length 301;
Best Local Similarity 44.4%; Pred. No. 42;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 NLVEQASTSQASLYQIYN 18
Db 64 NLIEQENRNKAGIYGIN 81

RESULT 10
O30XX5 PRELIMINARY; PRT; 418 AA.
AC O30XX5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Wilms' tumor suppressor 2a.
GN WT-T2A.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; UNKNOWN_1.
SQ SEQUENCE 110 AA; 12069 MW; 95FB6E170C7BECA4 CRC64;

Query Match 41.9%; Score 44; DB 6; Length 110;
Best Local Similarity 45.0%; Pred. No. 14;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
Db 91 IVEQCCASVCSLYQLSHYCN 110

RESULT 9
O33762 PRELIMINARY; PRT; 301 AA.
AC Q33762;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE ORf301 protein.
GN ORF301.
OS Allomyces macrogynus.
OG Mitochondrion.
OC Eukaryota; Fungi; Chytridiomycota; Blastocladales; Blastocladiaceae;
OC Allomyces.
OX NCBI_TaxID=28583;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96226032; PubMed=8636971;
RA Paquin B., Lang B.F.;
RT "The mitochondrial DNA of Allomyces macrogynus: the complete genomic
sequence from an ancestral fungus.";
RL J. Mol. Biol. 255:688-701(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Paquin B., Laforest M.J., Lang B.F.;
RT "80 conserved, palindromic sequence elements in the mitochondrial DNA
of Allomyces sp.: Evidence for mobility.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U01288; AAC49244.1;
DR InterPro; IPR003647; Intron_nuc_1.
DR InterPro; IPR003611; Intron_nuc_2.
DR InterPro; IPR000305; UvrC_N.
DR Pfam; PF01541; Excl_endo_N; 1.
DR SMART; SM00465; GYVC; 1.
DR SMART; SM00497; IENR1; 1.
DR SMART; SM00496; IENR2; 3.
KW Mitochondrion.
SQ SEQUENCE 301 AA; 34210 MW; F696732239DFCBAA CRC64;

Query Match 41.9%; Score 44; DB 8; Length 301;
Best Local Similarity 44.4%; Pred. No. 42;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 NLVEQASTSQASLYQIYN 18
Db 64 NLIEQENRNKAGIYGIN 81

RESULT 10
O30XX5 PRELIMINARY; PRT; 418 AA.
AC O30XX5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Wilms' tumor suppressor 2a.
GN WT-T2A.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
```

```
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21337607; PubMed=11444705;
RA Brunelli J.P., Robison B.D., Thorgaard G.H.;
RT "Ancient and recent duplications of the rainbow trout Wilms' tumor
gene.";
RL Genome 44:455-462(2001).
DR EMBL; AF334673; AAK52722.1; -.
DR InterPro; IPR000976; Wilms_tumour.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF02165; WT1; 1.
DR Pfam; PF00096; zf-C2H2; 4.
DR ProDom; PD000003; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_4.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 418 AA; 46116 MW; F1E9079668E9EED3 CRC64;

Query Match 41.9%; Score 44; DB 13; Length 418;
Best Local Similarity 55.6%; Pred. No. 59;
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 4 EQASTSQASLYQIYNFDN 21
Db 191 ENCPTSQALLLRNYSNDN 208

RESULT 11
O30XX4 PRELIMINARY; PRT; 418 AA.
ID Q30XX4;
AC Q30XX4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Wilms' tumor suppressor 2b.
GN WT-T2B.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21337607; PubMed=11444705;
RA Brunelli J.P., Robison B.D., Thorgaard G.H.;
RT "Ancient and recent duplications of the rainbow trout Wilms' tumor
gene.";
RL Genome 44:455-462(2001).
DR EMBL; AF334674; AAK52723.1; -.
DR InterPro; IPR000976; Wilms_tumour.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF02165; WT1; 1.
DR Pfam; PF00096; zf-C2H2; 4.
DR ProDom; PD000003; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_4.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 418 AA; 46102 MW; 85BAE6923ACF9DF6 CRC64;

Query Match 41.9%; Score 44; DB 13; Length 418;
Best Local Similarity 55.6%; Pred. No. 59;
Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 4 EQASTSQASLYQIYNFDN 21
Db 191 ENCPTSQALLLRNYSNDN 208

RESULT 12
O3Y5R9 PRELIMINARY; PRT; 512 AA.
ID O3Y5R9
AC O3Y5R9;
```



```

RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010798; AA004709.1; -
DR EMBL; AE010905; AA005584.1; -
KW Complete proteome.
SQ SSEQUENCE 553 AA; 64496 MW; 8470A3DB84048620 CRC64;

Query Match 41.9%; Score 44; DB 17; Length 553;
Best Local Similarity 42.1%; Pred. No. 80;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps

QY 2 LVEQASTSQASLYQIYNFD 20
:|: | | | | | | |
DB 530 IVEEMDTKQASLFSFLNME 548

RESULT 14
Q960S3 PRELIMINARY; PRT; 561 AA.
ID AC Q960S3;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE LD37788p.
DE GN BTZ OR G12878.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Chapeton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY051891; AAK93315.1; -
DR FlyBase; Fggn0045862; btz.
SQ SSEQUENCE 561 AA; 62974 MW; 57E75E2E0BF60D0 CRC64;

Query Match 41.9%; Score 44; DB 5; Length 561;
Best Local Similarity 52.9%; Pred. No. 81;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps

QY 4 EQASTSQASLYQIYNFD 20
:|: | | | | | | |
DB 22 EQAPKSRALLQTYGYD 38

RESULT 15
Q9URJ8 PRELIMINARY; PRT; 605 AA.
ID AC Q9URJ8;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE Glucose oxidase.
DE GN GOX.
OS Penicillium amagasakiense.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
OC NCBI_TaxID=63559;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33245;
RA Witt S., Kalisz H.M., Singh M.;
RT "Cloning, DNA sequencing and expression of the glucose oxidase gene
RT from Penicillium amagasakiense.";

```

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF012277; AAD01493.1; 1;
DR HSSP; P81156; 1GPE.
DR InterPro; IPR000172; GMC_Oxred.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00732; GMC_Oxred; 1; 1.
DR PROSITE; PS00623; GMC_Oxred_1; 1.
DR PROSITE; PS00624; GMC_Oxred_2; 1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN.1.
SQ SEQUENCE 605 AA; 65768 MW; F9F5E0A0687E8465 CRC64;

Query Match 41.9%; Score 44; DB 3; Length 605;
Best Local Similarity 52.6%; Pred. NO. 88;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 NLVEQASTSQASLYQIYNF 19
||| : : : : : : : :
Db 515 NLVONATLSQSDYVLQNF 533

RESULT 16

Q9VB04 PRELIMINARY; PRT; 761 AA.
AC Q9VB04;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CG12878 protein.
GN BTZ OR CGI2878.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila;
OX NCBI_TaxID=7227;
[1]

RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;

RA STRAIN-BERKELEY;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).
DR EMBL; AE003762; AAF56741.1; 1;
DR FlyBase; FBgn0045862; btz.
SQ SEQUENCE 761 AA; 83669 MW; A4ACCAA3E0F39204 CRC64;

Query Match 41.9%; Score 44; DB 5; Length 761;
Best Local Similarity 52.9%; Pred. NO. 11e+02;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 EQASTSQASLYQIYNF 20
||| : : : : : : : :
Db 222 EQAPKRSRAELLQTYGYD 238

RESULT 17

Q9NEK9 PRELIMINARY; PRT; 545 AA.
AC Q9NEK9; O18273;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ZC482.2 protein.
GN ZC482.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis;
OX NCBI_TaxID=6239;
[1]
RP SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
[3]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL132904; CAC35854.1; 1;
DR EMBL; Z93397; CAC35854.1; JOINED.
DR EMBL; Z93397; CAB07716.2; 1;
DR EMBL; AL132904; CAB07716.2; JOINED.
DR InterPro; IPR000494; EGFR_L domain.
DR Pfam; PF01030; Recep_L domain; 2.
SQ SEQUENCE 545 AA; 61178 MW; 8A64A9B337EB2276 CRC64;

Query Match 41.4%; Score 43.5; DB 5; Length 545;
Best Local Similarity 50.0%; Pred. NO. 95;
Matches 11; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

QY 3 VEOASTSQASLY---QIYNFDN 21
||| : : : : : : : :
Db 309 VEOISMIOYSTYKQYFNEN 330

RESULT 18

O10260 PRELIMINARY; PRT; 151 AA.
AC O10260;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Major capsid protein (Fragment).
OS Human papillomavirus.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10566;

[illegible]

Best Local Similarity 42.9%; Pred. No. 1.1e+02;
Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 NLVEQASTSOASLYQIYNFDN 21
DB 449 NLMEADAATAEISRSQIOWIN 469

RESULT 22

Q9RKU9 PRELIMINARY; PRT; 540 AA.

AC Q9RKU9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Malate synthase (EC 4.1.3.2)
GN ACEB OR SC06243 OR SCAH10.08C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=M130;
RX MEDLINE=20397121; PubMed=10941526;
RA Loke P., Slim T.S.;
RT "Molecular cloning, heterologous expression, and functional
characterisation of a malate synthase gene from Streptomyces
coelicolor A3(2).";
RL Can. J. Microbiol. 46:764-769(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1- PATHWAY: SECOND STEP IN GLYOXYLATE BYPASS, AN ALTERNATIVE TO THE
GLYOXYLATE.
CC -1- CATALYTIC ACTIVITY: L-MALATE + COA -> ACETYL-COA + H(2)O +
TRICARBOXYLIC ACID CYCLE (IN BACTERIA, FUNGI AND PLANTS).
CC -1- SIMILARITY: BELONGS TO THE MALATE SYNTHASE FAMILY.
DR EMBL; AL132824; CAB60162.1; -;
DR EMBL; AF206498; AAG29597.1; -;
DR InterPro; IPR001465; Malate_synthase.
DR Pfam; PF01274; Malate_synthase; 1.

DR PROSITE; PS00510; MALATE_SYNTHASE; 1.
KW Glyoxylate bypass; Lyase; Tricarboxylic acid cycle.
SQ SEQUENCE 540 AA; 59990 MW; 1CB229FBA611835 CRC64;

Query Match 41.0%; Score 43; DB 16; Length 540;
Best Local Similarity 42.9%; Pred. No. 1.1e+02;
Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 NLVEQASTSOASLYQIYNFDN 21
DB 449 NLMEADAATAEISRSQIOWIN 469

RESULT 23

Q9XJK1 PRELIMINARY; PRT; 714 AA.

AC Q9XJK1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative portal protein (Hypothetical 81.1 kDa protein).
GN L0114.
OS Bacteriophage 933W,
OS Bacteriophage VT2-Sa, and
OS Stx2 converting bacteriophage I.
OC Viruses; dsDNA viruses, no RNA stage; unclassified dsDNA phages.
OX NCBI_TaxID=10730, 97081, 180816;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Phage 933W;
RX MEDLINE=99173898; PubMed=10074068;
RA Plunkett G. III, Rose D.J., Durfee T.J., Blattner F.R.;
RT "Sequence of Shiga toxin 2 phage 933W from Escherichia coli O157:H7;
Shiga toxin as a phage late-gene product.";
RL J. Bacteriol. 181:1767-1778(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Phage 933W;
RA Plunkett G. III;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Phage VT2-Sa;
RA Miyamoto H.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Phage VT2-Sa;
RX MEDLINE=99419919; PubMed=10492170;
RA Miyamoto H., Nakai W., Yajima N., Fujibayashi A., Higuchi T., Sato K.,
RA Matsushiro A.;
RT "Sequence analysis of Stx2-Converting phage VT2-Sa shows a great
divergence in early regulation and replication regions.";
RL DNA Res. 6:235-240(1999).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=Stx2 converting bacteriophage I; STRAIN=STX2 PHAGE-I;
RA Sato T., Shimizu T., Watarai M., Kobayashi M., Kano S., Hamabata T.,
RA Yamasaki S., Takeda Y.;
RT "Genomic sequence of Shiga toxin 2-converting phage isolated from
Escherichia coli O157:H7 Okayama strain and comparison with other
Shiga toxin 2-converting phages.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF125520; AAD25457.1; -;
DR EMBL; AP000363; BAA84334.1; -;
DR EMBL; AP004402; BAB87850.1; -;
KW Hypothetical protein.
SQ SEQUENCE 714 AA; 81062 MW; 72D059AE7D103FE1 CRC64;

Query Match 41.0%; Score 43; DB 9; Length 714;
Best Local Similarity 47.4%; Pred. No. 1.5e+02;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:17 ; Search time 36.4706 Seconds
(without alignments)
109.609 Million cell updates/sec

Title: US-09-574-443-7_COPY_22_51

Perfect score: 170

Sequence: 1 DYNFHLGSHIREWLYVAGERGFNFDPKT 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	170	100.0	30	22 AAB48851	Mutant human insul
2	151	88.8	30	22 AAB48874	Mutant human insul
3	132	77.6	30	22 AAB48876	Mutant human insul
4	121	71.2	30	22 AAB48872	Mutant human insul
5	114	67.1	30	22 AAB48859	Mutant human insul
6	113	66.5	30	22 AAB48867	Mutant human insul
7	110	64.7	30	22 AAB48853	Mutant human insul
8	110	64.7	30	22 AAB48857	Mutant human insul
9	104	61.2	30	17 AAB88748	Asp(1) human insul
10	104	61.2	30	22 AAB48868	Mutant human insul

11	104	61.2	30	22 AAB48869	Mutant human insul
12	103	60.6	30	22 AAB48866	Mutant human insul
13	99	58.2	29	3 AAP20139	Sequence of des-Ph
14	99	58.2	29	6 AAP50834	Sequence of des-Ph
15	99	58.2	30	6 AAP50826	Sequence of human
16	99	58.2	30	7 AAP60909	Derivative of insu
17	99	58.2	30	7 AAP61335	Modified insulin B
18	99	58.2	30	11 AAR07755	Insulin analogue B
19	99	58.2	30	11 AAR04040	Amino acids 1-30 o
20	99	58.2	30	13 AAR30650	Phosphorylated hum
21	99	58.2	30	13 AAR26470	Sequence of human
22	99	58.2	30	14 AAR44454	Human insulin B-ch
23	99	58.2	30	15 AAR62642	Bovine and human i
24	99	58.2	30	15 AAR65881	Human insulin B-ch
25	99	58.2	30	15 AAR68896	Human insulin B-ch
26	99	58.2	30	17 AAR96000	Asymmetry sequence
27	99	58.2	30	17 AAR96002	Asymmetry sequence
28	99	58.2	30	19 AAW63755	Human insulin prot
29	99	58.2	30	20 AAO17828	Human insulin 2 fr
30	99	58.2	30	20 AAY15814	Amino acid sequenc
31	99	58.2	30	20 AAY08002	Human insulin B-ch
32	99	58.2	30	20 AAW93410	Human insulin deri
33	99	58.2	30	20 AAW81848	Human insulin beta
34	99	58.2	30	21 AAY51217	Human insulin B-ch
35	99	58.2	30	22 AAG65676	Fragment of an ins
36	99	58.2	30	22 AAG65689	Human proinsulin a
37	99	58.2	30	22 AAB74426	Human insulin B ch
38	99	58.2	30	22 AAB60422	Human insulin B-ch
39	99	58.2	30	22 AAB48849	Human insulin matu
40	99	58.2	30	22 AAB48861	Mutant human insul
41	99	58.2	30	22 AAB48870	Mutant human insul
42	99	58.2	30	23 AAU09050	Insulin/insulin-11
43	99	58.2	31	3 AAP20166	Insulin B-chain.
44	99	58.2	31	5 AAP40319	Sequence encoded b
45	99	58.2	31	5 AAP40810	Sequence encoded b

ALIGNMENTS

RESULT 1
AAB48851
ID AAB48851 standard; peptide; 30 AA.
AC AAB48851;
XX
XX 16-MAR-2001 (first entry)
DT
XX Mutant human insulin (IA protein) cysl B chain.
DE
DE Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW disulphide bond; cysteine replacement; mutant; muteln.
XX
XX Homo sapiens.
OS
OS Synthetic.
XX
XX WO200069901-A2.
PN
XX 23-NOV-2000.
PD
XX 19-MAY-2000; 2000WO-US13764.
PF
XX 19-MAY-1999; 99US-0134930.
PR
XX (XENC-) XENCOR INC.
PA
XX Dahiyat BI;
PI
XX WPI; 2001-025004/03.
DR
XX Non-naturally occurring protein with insulin activity useful for
XX treating type 1 and type 2 diabetes, comprising amino acid
PT

PT substitutions as compared to native human insulin and having enhanced
PT stability -
XX
PS Claim 10; Fig 3A; 95pp; English.
XX
CC The invention relates to novel non-naturally occurring mature human
CC insulin mutants, designated insulin activity (IA) proteins in the
CC specification (AAB48870-B48876), which have altered properties (e.g.,
CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
CC altered ability to oligomerise) compared to wild-type mature human
CC insulin (AAB48848, AAB48849) but which are still capable of binding to
CC an insulin receptor. The insulin mutants of the invention have less than
CC 98% sequence identity to wild-type human insulin, but have a 3D
CC structure which substantially corresponds to that of wild-type human
CC insulin. The invention also relates to recombinant nucleic acids encoding
CC the insulin mutants, expression vectors and host cells comprising mutant
CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
CC and a pharmaceutical composition comprising an insulin mutant of the
CC invention. The mature human insulin mutants are useful for treating
CC insulin-responsive conditions and disorders of carbohydrate metabolism
CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
CC insulin mutants are used in gene therapy techniques for treating these
CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin
CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.

XX Sequence 30 AA;
SQ Query Match 100.0%; Score 170; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 6e-17; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;

OY 1 DVNFHLYGSHIREWLYVAGERGFNFDPKT 30
DB 1 DVNFHLYGSHIREWLYVAGERGFNFDPKT 30

RESULT 2
AAB48874
ID AAB48874 standard; peptide; 30 AA.

XX AAB48874;
XX
XX 16-MAR-2001 (first entry)
XX Mutant human insulin (IA protein) trz_07 B chain.
XX
XX Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW mutant; mutein.

XX Homo sapiens.
OS Synthetic.

XX WO2000069901-A2.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000WO-US13764.

XX 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.

XX Dahiyat BI;

XX WPI; 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for
PT treating type 1 and type 2 diabetes, comprising amino acid

PT substitutions as compared to native human insulin and having enhanced
PT stability -

XX Claim 10; Fig 5B; 95pp; English.

XX
CC The invention relates to novel non-naturally occurring mature human
CC insulin mutants, designated insulin activity (IA) proteins in the
CC specification (AAB48850-B48876), which have altered properties (e.g.,
CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
CC altered ability to oligomerise) compared to wild-type mature human
CC insulin (AAB48848, AAB48849) but which are still capable of binding to
CC an insulin receptor. The insulin mutants of the invention have less than
CC 98% sequence identity to wild-type human insulin, but have a 3D
CC structure which substantially corresponds to that of wild-type human
CC insulin. The invention also relates to recombinant nucleic acids encoding
CC the insulin mutants, expression vectors and host cells comprising mutant
CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
CC and a pharmaceutical composition comprising an insulin mutant of the
CC invention. The mature human insulin mutants are useful for treating
CC insulin-responsive conditions and disorders of carbohydrate metabolism
CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
CC insulin mutants are used in gene therapy techniques for treating these
CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin
CC hexamers. Sequences AAB48871-B48876 represent the A and B chains of
CC insulin mutants which have improved stability.

XX Sequence 30 AA;

SQ Query Match 88.8%; Score 151; DB 22; Length 30;
Best Local Similarity 90.0%; Pred. No. 2.9e-14;
Matches 27; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 DVNFHLYGSHIREWLYVAGERGFNFDPKT 30
DB 1 DVNFHLYGSHIREWLYVAGERGFNFDPKT 30

RESULT 3

AAB48876
ID AAB48876 standard; peptide; 30 AA.

XX AAB48876;

XX 16-MAR-2001 (first entry)

XX Mutant human insulin (IA protein) trz_08 B chain.

XX Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW mutant; mutein.

XX Homo sapiens.
OS Synthetic.

XX WO2000069901-A2.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000WO-US13764.

XX 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.

XX Dahiyat BI;

XX WPI; 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for
PT treating type 1 and type 2 diabetes, comprising amino acid
PT substitutions as compared to native human insulin and having enhanced
PT stability -

XX PS Claim 10; Fig 5C; 95pp; English.

XX CC The invention relates to novel non-naturally occurring mature human

XX CC insulin mutants, designated insulin activity (IA) proteins in the

XX CC specification (AAB48850-B48876), which have altered properties (e.g.,

XX CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,

XX CC altered ability to oligomerise) compared to wild-type mature human

XX CC insulin (AAB48848, AAB48849) but which are still capable of binding to

XX CC an insulin receptor. The insulin mutants of the invention have less than

XX CC 98% sequence identity to wild-type human insulin, but have a 3D

XX CC structure which substantially corresponds to that of wild-type human

XX CC insulin. The invention also relates to recombinant nucleic acids encoding

XX CC the insulin mutants, expression vectors and host cells comprising mutant

XX CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,

XX CC and a pharmaceutical composition comprising an insulin mutant, of the

XX CC invention. The mature human insulin mutants are useful for treating

XX CC insulin-responsive conditions and disorders of carbohydrate metabolism

XX CC such as type 1 or type 2 diabetes. The nucleic acids encoding the

XX CC insulin mutants are used in gene therapy techniques for treating these

XX CC conditions. The insulin mutants are more stable compared to wild-type

XX CC mature human insulin, or have an improved ability to form insulin

XX CC hexamers. Sequences AAB48871-B48876 represent the A and B chains of

XX CC insulin mutants which have improved stability.

XX SQ Sequence 30 AA;

Query Match 77.6%; Score 132; DB 22; Length 30;

Best Local Similarity 83.3%; Pred. No. 1.4e-11;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 DVNFHLYGSHIREWLVLVAGERGFNFDPKT 30

DB 1 DKNFHLCLSHIREWLVLVCCGERGFNFDNKT 30

RESULT 4

AAB48872

ID AAB48872 standard; peptide; 30 AA.

AC AAB48872;

XX 16-MAR-2001 (first entry)

DE Mutant human insulin (IA protein) trz_06 B chain.

XX Human insulin; insulin activity protein; IA protein;

KW diabetes mellitus; type 1; type 2; stability;

KW mutant; mutein.

OS Homo sapiens.

OS Synthetic.

XX WO200069901-A2.

PN 23-NOV-2000.

PD 19-MAY-2000; 2000WO-US13764.

XX 19-MAY-1999; 99US-0134930.

PR (XENC-) XENCOR INC.

PA Dahiyat BI;

XX WPI; 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for

PT treating type 1 and type 2 diabetes, comprising amino acid

PT substitutions as compared to native human insulin and having enhanced

PT stability -

XX Claim 10; Fig 5A; 95pp; English.

XX CC The invention relates to novel non-naturally occurring mature human

XX CC insulin mutants, designated insulin activity (IA) proteins in the

XX CC specification (AAB48850-B48876), which have altered properties (e.g.,

XX CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,

XX CC altered ability to oligomerise) compared to wild-type mature human

XX CC insulin (AAB48848, AAB48849) but which are still capable of binding to

XX CC an insulin receptor. The insulin mutants of the invention have less than

XX CC 98% sequence identity to wild-type human insulin, but have a 3D

XX CC structure which substantially corresponds to that of wild-type human

XX CC insulin. The invention also relates to recombinant nucleic acids encoding

XX CC the insulin mutants, expression vectors and host cells comprising mutant

XX CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,

XX CC and a pharmaceutical composition comprising an insulin mutant, of the

XX CC invention. The mature human insulin mutants are useful for treating

XX CC insulin-responsive conditions and disorders of carbohydrate metabolism

XX CC such as type 1 or type 2 diabetes. The nucleic acids encoding the

XX CC insulin mutants are used in gene therapy techniques for treating these

XX CC conditions. The insulin mutants are more stable compared to wild-type

XX CC mature human insulin, or have an improved ability to form insulin

XX CC hexamers. Sequences AAB48871-B48876 represent the A and B chains of

XX CC insulin mutants which have improved stability.

XX SQ Sequence 30 AA;

Query Match 71.2%; Score 121; DB 22; Length 30;

Best Local Similarity 76.7%; Pred. No. 4.8e-10;

Matches 23; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 DVNFHLYGSHIREWLVLVAGERGFNFDPKT 30

DB 1 DVNQHLGCSHLVREALYLVCCGERGFNFDPKT 30

RESULT 5

AAB48859

ID AAB48859 standard; peptide; 30 AA.

AC AAB48859;

XX 16-MAR-2001 (first entry)

DE Mutant human insulin (IA protein) cysd+ B chain.

XX Human insulin; insulin activity protein; IA protein;

KW diabetes mellitus; type 1; type 2; stability;

KW disulphide bond; cysteine replacement; mutant; mutein.

OS Homo sapiens.

OS Synthetic.

XX WO200069901-A2.

PN 23-NOV-2000.

PD 19-MAY-2000; 2000WO-US13764.

XX 19-MAY-1999; 99US-0134930.

PR (XENC-) XENCOR INC.

PA Dahiyat BI;

XX WPI; 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for

PT treating type 1 and type 2 diabetes, comprising amino acid

PT substitutions as compared to native human insulin and having enhanced

PT stability -

XX Claim 10; Fig 3E; 95pp; English.

XX The invention relates to novel non-naturally occurring mature human

insulin mutants, designated insulin activity (IA) proteins in the specification (AAB48850-B48876), which have altered properties (e.g., enhanced oxidative, alkaline or thermal stability, prolonged shelf-life, altered ability to oligomerise) compared to wild-type mature human insulin (AAB48848, AAB48849) but which are still capable of binding to an insulin receptor. The insulin mutants of the invention have less than 98% sequence identity to wild-type human insulin, but have a 3D structure which substantially corresponds to that of wild-type human insulin. The invention also relates to recombinant nucleic acids encoding the insulin mutants, expression vectors and host cells comprising mutant insulin-encoding DNA, the recombinant preparation of an insulin mutant, and a pharmaceutical composition comprising an insulin mutant of the invention. The mature human insulin mutants are useful for treating insulin-responsive conditions and disorders of carbohydrate metabolism such as type 1 or type 2 diabetes. The nucleic acids encoding the insulin mutants are used in gene therapy techniques for treating these conditions. The insulin mutants are more stable compared to wild-type mature human insulin, or have an improved ability to form insulin hexamers. Sequences AAB48850-B48863 represent the A and B chains of insulin mutants in which one or more cysteine residue has been replaced, thus preventing the formation of at least one disulphide bond and improving stability and activity.

XX SQ Sequence 30 AA;

Query Match 67.1%; Score 114; DB 22; Length 30;
Best Local Similarity 72.4%; Pred. No. 4.7e-09;
Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVACGERGFNFDPKT 30
||:|||||:| ||| ||||| : |||
DB 2 VNFHLYGSHLVVEALVLCGERGFFFTPKT 30

RESULT 6

AAB48867
ID AAB48867 standard; peptide; 30 AA.
XX
AC AAB48867;
XX
DT 16-MAR-2001 (first entry)
XX
DE Mutant human insulin B chain A14W.
XX
KW Human insulin; insulin activity protein; IA protein; B chain;
KW diabetes mellitus; type 1; type 2; stability;
KW oligomerisation; hexamer formation; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200069901-A2.
XX
PD 23-NOV-2000.
XX
PF 19-MAY-2000; 2000WO-US13764.
XX
PR 19-MAY-1999; 99US-0134930.
XX
PA (XENC-) XENCOR INC.
XX
PI Dahiyat BI;
XX
DR WPI; 2001-025004/03.
XX
PT Non-naturally occurring protein with insulin activity useful for
PT treating type 1 and type 2 diabetes, comprising amino acid
PT substitutions as compared to native human insulin and having enhanced
PT stability -
XX
PS Claim 10; Fig 4D; 95pp; English.
XX
CC The invention relates to novel non-naturally occurring mature human

insulin mutants, designated insulin activity (IA) proteins in the specification (AAB48850-B48876), which have altered properties (e.g., enhanced oxidative, alkaline or thermal stability, prolonged shelf-life, altered ability to oligomerise) compared to wild-type mature human insulin (AAB48848, AAB48849) but which are still capable of binding to an insulin receptor. The insulin mutants of the invention have less than 98% sequence identity to wild-type human insulin, but have a 3D structure which substantially corresponds to that of wild-type human insulin. The invention also relates to recombinant nucleic acids encoding the insulin mutants, expression vectors and host cells comprising mutant insulin-encoding DNA, the recombinant preparation of an insulin mutant, and a pharmaceutical composition comprising an insulin mutant of the invention. The mature human insulin mutants are useful for treating insulin-responsive conditions and disorders of carbohydrate metabolism such as type 1 or type 2 diabetes. The nucleic acids encoding the insulin mutants are used in gene therapy techniques for treating these conditions. The insulin mutants are more stable compared to wild-type mature human insulin, or have an improved ability to form insulin hexamers. Sequences AAB48864-B48870 represent the mutant B chains of insulin mutants which have an improved ability to form hexamers. The A chain used in these insulin mutants is identical to the wild-type mature human insulin A chain (AAB48850).

XX SQ Sequence 30 AA;

Query Match 66.5%; Score 113; DB 22; Length 30;
Best Local Similarity 72.4%; Pred. No. 6.5e-09;
Matches 21; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVACGERGFNFDPKT 30
||:|||||:| ||| ||||| : |||
DB 2 VNFHLYGSHLVVEALVLCGERGFFFTPKT 30

RESULT 7

AAB48853
ID AAB48853 standard; peptide; 30 AA.
XX
AC AAB48853;
XX
DT 16-MAR-2001 (first entry)
XX
DE Mutant human insulin (IA protein) cys77a B chain.
XX
KW Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW disulphide bond; cysteine replacement; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200069901-A2.
XX
PD 23-NOV-2000.
XX
PF 19-MAY-2000; 2000WO-US13764.
XX
PR 19-MAY-1999; 99US-0134930.
XX
PA (XENC-) XENCOR INC.
XX
PI Dahiyat BI;
XX
DR WPI; 2001-025004/03.
XX
PT Non-naturally occurring protein with insulin activity useful for
PT treating type 1 and type 2 diabetes, comprising amino acid
PT substitutions as compared to native human insulin and having enhanced
PT stability -
XX
PS Claim 10; Fig 3B; 95pp; English.
XX
CC The invention relates to novel non-naturally occurring mature human

insulin mutants, designated insulin activity (IA) proteins in the specification (AAB48850-B48876), which have altered properties (e.g., enhanced oxidative, alkaline or thermal stability, prolonged shelf-life, altered ability to oligomerise) compared to wild-type mature human insulin (AAB48848, AAB48849) but which are still capable of binding to an insulin receptor. The insulin mutants of the invention have less than 98% sequence identity to wild-type human insulin, but have a 3D structure which substantially corresponds to that of wild-type human insulin. The invention also relates to recombinant nucleic acids encoding the insulin mutants, expression vectors and host cells comprising mutant, insulin-encoding DNA, the recombinant preparation of an insulin mutant, and a pharmaceutical composition comprising an insulin mutant of the invention. The mature human insulin mutants are useful for treating insulin-responsive conditions and disorders of carbohydrate metabolism such as type 1 or type 2 diabetes. The nucleic acids encoding the insulin mutants are used in gene therapy techniques for treating these conditions. The insulin mutants are more stable compared to wild-type mature human insulin, or have an improved ability to form insulin hexamers. Sequences AAB48850-B48863 represent the A and B chains of insulin mutants in which one or more cysteine residue has been replaced, thus preventing the formation of at least one disulphide bond and improving stability and activity.

Sequence 30 AA;

Query Match 64.7%; Score 110; DB 22; Length 30;
Best Local Similarity 71.4%; Pred. No. 1.7e-08;
Matches 20; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 NFHLGSHIREWLYLVAGERGFNDPKT 30
:|||||: | |||| | |||| : |||
DB 3 NYHLGSHLVEALYLVCGERGFFYTPKT 30

RESULT 8

AAB48857
ID AAB48857 standard; peptide; 30 AA.

AC AAB48857;

DT 16-MAR-2001 (first entry)

DE Mutant human insulin (IA protein) cys77d B chain.

KW Human insulin; insulin activity protein; IA protein;
diabetes mellitus; type 1; type 2; stability;
disulphide bond; cysteine replacement; mutant; mutein.

OS Homo sapiens.

OS Synthetic.

PN WO200069901-A2.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000WO-US13764.

XX 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.

XX Dahiyat BI;

XX WPI; 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for
PT treating type 1 and type 2 diabetes, comprising amino acid
PT substitutions as compared to native human insulin and having enhanced
PT stability.

XX Claim 10; Fig 3d; 95pp; English.

XX The invention relates to novel non-naturally occurring mature human

insulin mutants, designated insulin activity (IA) proteins in the specification (AAB48850-B48876), which have altered properties (e.g., enhanced oxidative, alkaline or thermal stability, prolonged shelf-life, altered ability to oligomerise) compared to wild-type mature human insulin (AAB48848, AAB48849) but which are still capable of binding to an insulin receptor. The insulin mutants of the invention have less than 98% sequence identity to wild-type human insulin, but have a 3D structure which substantially corresponds to that of wild-type human insulin. The invention also relates to recombinant nucleic acids encoding the insulin mutants, expression vectors and host cells comprising mutant, insulin-encoding DNA, the recombinant preparation of an insulin mutant, and a pharmaceutical composition comprising an insulin mutant of the invention. The mature human insulin mutants are useful for treating insulin-responsive conditions and disorders of carbohydrate metabolism such as type 1 or type 2 diabetes. The nucleic acids encoding the insulin mutants are used in gene therapy techniques for treating these conditions. The insulin mutants are more stable compared to wild-type mature human insulin, or have an improved ability to form insulin hexamers. Sequences AAB48850-B48863 represent the A and B chains of insulin mutants in which one or more cysteine residue has been replaced, thus preventing the formation of at least one disulphide bond and improving stability and activity.

Sequence 30 AA;

Query Match 64.7%; Score 110; DB 22; Length 30;
Best Local Similarity 71.4%; Pred. No. 1.7e-08;
Matches 20; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 NFHLGSHIREWLYLVAGERGFNDPKT 30
:|||||: | |||| | |||| : |||
DB 3 NYHLGSHLVEALYLVCGERGFFYTPKT 30

RESULT 9

AAR88748
ID AAR88748 standard; peptide; 30 AA.

AC AAR88748;

DT 20-SEP-1996 (first entry)

XX Asp(1) human insulin B-chain.

XX hyperglycaemia; diabetes mellitus; zinc; hexamer.

XX Synthetic.

XX Key Location/Qualifiers

FT Cross-links 7

FT /note= "disulphide-bonded with Cys(7) of the A-Chain"

FT Cross-links 19

FT /note= "disulphide-bonded with Cys(20) of the A-Chain"

FT Misc-difference 13

FT /label= Glu, Gln

XX WO9604307-A1.

XX 15-FEB-1996.

XX 01-AUG-1995; 95WO-US09768.

XX 02-AUG-1994; 94US-0285661.

XX (ELIL) LILLY & CO ELI.

XX Bakaysa DL, Brems DN;

XX WPI; 1996-129336/13.

XX New human insulin analogues contg. B1 aspartic acid - and opt. B13
PT glutamine, have increased tendency to self associate, providing slow
PT onset of activity but long lasting effect

XX
PS Claim 1; Page 38; 49pp; English.

A new human insulin analogue is provided in which the B chain has Asp at position 1 and optionally Gln instead of Glu at position 13. Also provided are hexamers consisting of six of the new analogues and from 0 to 2 zinc ions associated with each hexamer. The presence of Asp(B1) results in the stable formation of an allosteric Zn-hexamer configuration (normally induced only by ligands); and the presence of Gln(B13) causes self-association of the analogue to an R-state conformation in the absence of Zn. The analogues are useful for treating hyperglycaemia. They have a slow onset of activity but long-lasting action. They are soluble (avoiding complications involved in suspension formulations) and have increased tendency to self-associate.

SQ	Sequence	30 AA;
Query Match	61.2%;	Score 104; DB 17; Length 30;
Best Local Similarity	66.7%;	Pred. No. 1.2e-07;
Matches	20; Conservative	3; Mismatches 7; Indels 0; Gaps 0;

QY 1 DVNFHLYGSHIREWLYLVAGERGFNFDPKT 30
 IIII IIII : IIII IIII : IIII
 Db 1 DVNOHLCGSHLVZALYLVCGERGEFFYTPKT 30

RESULT 10
AAB48868
ID AAB48868 standard; peptide: 30 AA.

XX	ARB48868;
XX	
XX	16-MAR-2001 (first entry)
DT	
XX	
DE	Mutant human insulin B chain H5F/A14W

XX Human insulin; insulin activity protein; IA protein; B chain;
 KW diabetes mellitus; type 1; type 2; stability;
 KW oligomerisation; hexamer formation; mutant; mutagen.
 KW

AA	Homo sapiens.
OS	Synthetic.
OS	WO200069901-A2.
XX	23-NOV-2000.
PN	
XX	
PD	

XX 19-MAY-2000; 2000WO-US13764.
PF
XX
PR 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.
XX PA
XX XX
XX PI Dahiyat BI;
XX XX
DR WPI: 2001-025004/03

Non-naturally occurring protein with insulin activity useful for treating type 1 and type 2 diabetes, comprising amino acid substitutions as compared to native human insulin and having enhanced stability -

PS Claim 10; Fig 4E; 95pp; English.

The invention relates to novel non-naturally occurring mature human insulin mutants, designated insulin activity (IA) proteins in the specification (AA948850-948876), which have altered properties (e.g., enhanced oxidative, alkaline or thermal stability, prolonged shelf-life, altered ability to oligomerise) compared to wild-type mature human insulin (AA948848, AA948849) but which are still capable of binding to an insulin receptor. The insulin mutants of the invention have less than 98% sequence identity to wild-type human insulin, but have a 3D

structure which substantially corresponds to that of wild-type human insulin. The invention also relates to recombinant nucleic acids encoding the insulin mutants, expression vectors and host cells comprising mutant insulin-encoding DNA, the recombinant preparation of an insulin mutant, and a pharmaceutical composition comprising an insulin mutant of the invention. The mature human insulin mutants are useful for treating insulin-responsive conditions and disorders of carbohydrate metabolism such as type 1 or type 2 diabetes. The nucleic acids encoding the insulin mutants are used in gene therapy techniques for treating these conditions. The insulin mutants are more stable compared to wild-type mature human insulin, or have an improved ability to form insulin hexamers. Sequences AA84864-848870 represent the mutant B chains of insulin mutants which have an improved ability to form hexamers. The A chain used in these insulin mutants is identical to the wild-type mature human insulin A chain (AA848850).

SQ	Sequence	30 AA;
Query Match	61.2%;	Score 104; DB 22; Length 30;
Best Local Similarity	69.0%;	Pred. No. 1.2e-07;
Matches 20; Conservative	2; Mismatches	7; Indels 0; Gaps 0

QY 2 VNFHLYGSHIREWLYLVAGERGFNFDPKT 30
 11 1 111: 11111 11111 : 111

Db 2 VNOFLCGSHLVEWLYLVCGERGFYTPKT 30

RESULT 11
AAB48869
ID AAB48869 standard; peptide: 30 AA.

XX	AB48869;
AC	
XX	16-MAR-2001 (first entry)
DT	
XX	Mutant human insulin B chain A14Y.
DE	

XX Human insulin; insulin activity protein; IA protein; B chain;
KW diabetes mellitus; type 1; type 2; stability;
KW oligomerisation; hexamer formation; mutant; mutn.
KW

XX	Homo sapiens.
OS	Synthetic.
OS	
XX	WO200069901-A2.
PN	
XX	23-NOV-2000
PD	

XX
PF 19-MAY-2000; 2000WO-US13764.
XX
PR 19-MAY-1999; 99US-0134930.

XX. (XENC-) XENCOR INC.
PA
XX
XX
PI Dahiyat BI;
XX
DR WPT: 2001-025004/03

Non-naturally occurring protein with insulin activity useful for treating type 1 and type 2 diabetes, comprising amino acid substitutions as compared to native human insulin and having enhanced stability -

XX
PS Claim 10; Fig. 4F; 95pp; English.

The invention relates to novel non-naturally occurring mature human insulin mutants, designated insulin activity (IA) proteins in the specification (AAB48850-B48876), which have altered properties (e.g., enhanced oxidative, alkaline or thermal stability, prolonged shelf-life, altered ability to oligomerise) compared to wild-type mature human insulin (AAB48848, AAB48849) but which are still capable of binding to an insulin receptor. The insulin mutants of the invention have less than 98% sequence identity to wild-type human insulin, but have a 3D

CC structure which substantially corresponds to that of wild-type human
 CC insulin. The invention also relates to recombinant nucleic acids encoding
 CC the insulin mutants, expression vectors and host cells comprising mutant
 CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
 CC and a pharmaceutical composition comprising an insulin mutant of the
 CC invention. The mature human insulin mutants are useful for treating
 CC insulin-responsive conditions and disorders of carbohydrate metabolism
 CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
 CC insulin mutants are used in gene therapy techniques for treating these
 CC conditions. The insulin mutants are more stable compared to wild-type
 CC mature human insulin, or have an improved ability to form insulin
 CC hexamers. Sequences AAB48864-B48870 represent the mutant B chains of
 CC insulin mutants which have an improved ability to form hexamers.
 CC The A chain used in these insulin mutants is identical to the wild-type
 CC mature human insulin A chain (AAB48850).

XX Sequence 30 AA;

Query Match 61.2%; Score 104; DB 22; Length 30;
 Best Local Similarity 69.0%; Pred. NO. 1.2e-07;
 Matches 20; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
 II III IIII : IIII : IIII
 DB 2 VNHLCGSHLVEYLYLVCGRGFFTYPTKT 30

RESULT 12

AAB48866
 ID AAB48866 standard; peptide; 30 AA.

XX AC AAB48866;

XX DT 16-MAR-2001 (first entry)

XX DE Mutant human insulin B chain A14F.

XX DE Human insulin; insulin activity protein; IA protein; B chain;
 KW diabetes mellitus; type 1; type 2; stability;
 KW oligomerisation; hexamer formation; mutant; mutein.

XX OS Homo sapiens.
 OS Synthetic.

XX PN WO200069901-A2.

XX PD 23-NOV-2000.

XX PF 19-MAY-2000; 2000WO-US13764.

XX PR 19-MAY-1999; 99US-0134930.

XX PA (XENC-) XENCOR INC.

XX PI Dahiyat BI;

XX DR WPI; 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for
 PT treating type 1 and type 2 diabetes, comprising amino acid
 PT substitutions as compared to native human insulin and having enhanced
 PT stability -

XX Claim 10; Fig 4C; 95pp; English.

XX The invention relates to novel non-naturally occurring mature human
 CC insulin mutants, designated insulin activity (IA) proteins in the
 CC specification (AAB48850-B48870), which have altered properties (e.g.,
 CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
 CC altered ability to oligomerise) compared to wild-type mature human
 CC insulin (AAB48848, AAB48849) but which are still capable of binding to
 CC an insulin receptor. The insulin mutants of the invention have less than
 CC 98% sequence identity to wild-type human insulin, but have a 3D

CC structure which substantially corresponds to that of wild-type human
 CC insulin. The invention also relates to recombinant nucleic acids encoding
 CC the insulin mutants, expression vectors and host cells comprising mutant
 CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
 CC and a pharmaceutical composition comprising an insulin mutant of the
 CC invention. The mature human insulin mutants are useful for treating
 CC insulin-responsive conditions and disorders of carbohydrate metabolism
 CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
 CC insulin mutants are used in gene therapy techniques for treating these
 CC conditions. The insulin mutants are more stable compared to wild-type
 CC mature human insulin, or have an improved ability to form insulin
 CC hexamers. Sequences AAB48864-B48870 represent the mutant B chains of
 CC insulin mutants which have an improved ability to form hexamers.
 CC The A chain used in these insulin mutants is identical to the wild-type
 CC mature human insulin A chain (AAB48850).

XX Sequence 30 AA;

Query Match 60.6%; Score 103; DB 22; Length 30;
 Best Local Similarity 69.0%; Pred. NO. 1.7e-07;
 Matches 20; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
 II III IIII : IIII : IIII
 DB 2 VNHLCGSHLVEYLYLVCGRGFFTYPTKT 30

RESULT 13

AAP20139
 ID AAP20139 standard; Protein; 29 AA.

XX AC AAP20139;

XX DT 19-AUG-1992 (first entry)

XX DE Sequence of des-Phe(B1) (human) insulin analogue chain B2.

XX KW Diabetes; therapy; insulin; hormone.

XX OS Homo sapiens.

XX Key Location/Qualifiers
 FH Disulfide-bond 6

FT /note= "bonds to Cys(7) of the A1 chain"

FT Disulfide-bond 18

FT /note= "bonds to Cys(20) of the A1 chain"

FT Misc-difference 30

FT /label= Thr, OH,

XX EP46979-A.

XX PD 10-MAR-1982.

XX PF 26-AUG-1981; 81EP-0106625.

XX PR 03-SEP-1980; 80DE-3033127.

XX PA (FARH) HOECHST AG.

XX PI Geiger R;

XX DR WPI; 1982-20140E/11 (20140E).

XX Antidiabetic des-B1-phenylalanine human insulin - prepd. e.g. by
 PT Edman degradation of human insulin

XX PS Claim 1; Page 9; 15pp; German.

XX The insulin analogues of the invention induce the formation of
 CC antibodies only to a slight extent, and show better glucose
 CC tolerance than porcine insulin. When residue 21 of chain A is Thr,
 CC the analogue has high solubility and is therefore suitable for the
 CC prodn. of conc. solns. for insulin pumps.

```
XX SQ Sequence 29 AA;
Query Match 58.2%; Score 99; DB 3; Length 29;
Best Local Similarity 69.0%; Pred. No. 5.8e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGRGFNFDPKT 30
DB 1 VNQHLCGSHLVEALYLVCGERGFFYTPKT 29

RESULT 14
AAP50834
ID AAP50834 standard; protein; 29 AA.
XX AC AAP50834;
XX DT 28-NOV-1991 (first entry)
XX DE Sequence of des-PheB1- human insulin -(B30)-OBut.
XX KW Hormone; insulin; diabetes; therapy; sustained release.
XX OS Homo sapiens.
XX PN DE3334407-A.
XX PD 04-APR-1985.
XX PF 23-SEP-1983; 83DE-3334407.
XX PR 23-SEP-1983; 83DE-3334407.
XX PA (FARH ) HOECHST AG.
XX PI Grau U, Geiger R, Obermeier R;
XX DR WPI; 1985-087999/15.
XX PT New insulin derivs. with modified B30 position - having long
XX PT lasting antidiabetic activity
XX PS Disclosure; Page 14; 28pp; German.
XX CC The inventors claim new insulin derivs. with modified B30 position.
XX CC When used as antidiabetics, the insulin derivs. match the natural
XX CC hormone balance more closely than normal insulin formulations.
XX CC Because of their inherently low solubility at physiological pH, they
XX CC provide sustained release formulations without requiring additives
XX CC such as zinc. They are about as active as insulin (on a molar basis)
XX CC and no more immunogenic. See also DE3327928, DE3327709 and
XX CC DE3326473.
XX SQ Sequence 29 AA;

Query Match 58.2%; Score 99; DB 6; Length 29;
Best Local Similarity 69.0%; Pred. No. 5.8e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGRGFNFDPKT 30
DB 1 VNQHLCGSHLVEALYLVCGERGFFYTPKT 29

RESULT 15
AAP50826
ID AAP50826 standard; protein; 30 AA.
XX AC AAP50826;
XX DT 28-NOV-1991 (first entry)
XX
```

```
DE DE Sequence of human insulin B-chain deriv.
DE DE Des-Phe-B1-h-insulin-ArgB31.
XX KW Insulin derivative; crystal suspension; diabetes; therapy.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Modified-site 31
FT /label= Arg-OH
XX PN DE3327709-A.
XX PD 07-FEB-1985.
XX PF 29-JUL-1983; 83DE-3327709.
XX PR 29-JUL-1983; 83DE-3327709.
XX PA (FARH ) HOECHST AG.
XX PI Grau U;
XX DR WPI; 1985-039026/07.
XX CC Insulin derivs. crystal suspensions prodn. - by crystallisation
XX CC in aq. medium at isoelectric pt. in presence of aromatic hydroxy
XX CC cpds.
XX PS Disclosure; Page 10; 21pp; German.
XX CC The inventors claim a method for the prodn. of crystal suspensions
XX CC of one or more insulin derivs. having an isoelectric pt. of 5.8 -
XX CC 8.5. The insulin derivs. crystallise in the form of very uniform
XX CC prisms ca. 10 microns in size giving predictable and reproducible
XX CC dissolution characteristics. The crystal suspensions can have
XX CC relatively low zinc contents, so that they can be freely mixed with
XX CC insulin soln. without causing pptn. of the dissolved insulin.
XX CC See also DE3327928, DE3334407 and DE3326473.
XX SQ Sequence 30 AA;

Query Match 58.2%; Score 99; DB 6; Length 30;
Best Local Similarity 69.0%; Pred. No. 6.1e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGRGFNFDPKT 30
DB 1 VNQHLCGSHLVEALYLVCGERGFFYTPKT 29

RESULT 16
AAP60909
ID AAP60909 standard; Protein; 30 AA.
XX AC AAP60909;
XX DT 26-JUN-1991 (first entry)
XX DE Derivative of insulin A-chain.
XX KW Insulin; hormone.
XX PN EP194864-A.
XX PD 17-SEP-1986.
XX PF 11-MAR-1986; 86EP-0301755.
XX PR 12-MAR-1985; 85DK-0001135.
XX PA (NOVO ) NOVO INDUSTRI A/S.
XX
```

PS	Claim 4; Page 19; 18pp; English.
XX	Modified insulin B-chain has one or more protected groups which act
CC	to protract the effective life of the hormone, varying in degree
CC	according to blocking groups.
XX	
SQ	Sequence 30 AA;
	Query Match 58.2%; Score 99; DB 7; Length 30;
	Best Local Similarity 69.0%; Pred. No. 6.le-07;
	Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY	2 VNFHLYGSHIREWLYLVAGERGFNFDPKT 30 : :
Dd	2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30 : :
RESULT 18	
AAR07755	ID AAR07755 standard; protein; 30 AA.
XX	AC AAR07755;
XX	DT 14-FEB-1991 (first entry)
XX	DE Insulin analogue.B-chain.
XX	KW insulins analogue; diabetes; liver.
OS	Synthetic.
XX	Key Location/Qualifiers
FH	Disulfide-bond 7..7
FT	/note="bonded to Cys(7) of A-chain"
FT	Disulfide-bond 19..19
FT	/note="bonded to Cys(20) of A-chain"
XX	WO9012814-A.
PN	01-NOV-1990.
PD	17-APR-1990; 9OWO-US02070.
Pf	20-APR-1989; 89US-0340929.
PR	(MOUN) MOUNT SINAI SCHOOL.
PA	Katsoyannis PG;
PI	WPI; 1990-348431/46.
XX	New hepato-specific insulin analogues - contain tryptophan or
XX	other bulky, hydrophobic residues substd. at A14 and A19 and opt.
PT	A13, A15 and B16 positions of insulin polypeptide
PT	Claim 5; Page 27; 36pp; English.
XX	The analogue has Trp substituted at position 14 of the A chain. This
XX	renders the analogue hepatospecific, being more active in the liver
CC	than in the adipose tissue or peripheral tissue. It closely mimics
CC	the metabolic balance between the liver and peripheral tissues.
CC	See also AAR07754 and AAR08364.
XX	
SQ	Sequence 30 AA;
	Query Match 58.2%; Score 99; DB 11; Length 30;
	Best Local Similarity 69.0%; Pred. No. 6.le-07;
	Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY	2 VNFHLYGSHIREWLYLVAGERGFNFDPKT 30 : :
Dd	2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30 : :

RESULT 19
 AAR04040
 ID AAR04040 standard; protein; 30 AA.
 AC AAR04040;
 XX
 DT 30-MAY-1989 (first entry)
 XX
 DE Amino acids 1-30 of human insulin B chain.
 XX
 KW Mini-pro-insulin; insulin; fusion protein; diabetes mellitus.
 XX
 FH Key Location/Qualifiers
 FT CDS 15..102
 PN EP347781-A.
 XX
 PD 27-DEC-1989.
 XX
 PF 17-JUN-1989; 89EP-0111027.
 XX
 PR 23-JUN-1988; 88DE-3821159.
 XX
 PA (FARH) HOESCHT AG.
 XX
 PI Dorschug M, Habermann P, Seipke G, Uhlmann E;
 DR WPI; 1990-001166/01.
 DR N-PSDB; AAQ02824.
 XX
 PT Mini-pro-insulin contg. A and B chains joined by arginine - useful for
 PT treating diabetes and as intermediate for human insulin, and new DNA,
 PT fusion proteins, etc.
 XX
 PS Disclosure; Page 11; 21pp; English.
 XX
 CC This sequence of amino acids (AA's) 1-30 of human insulin B chain is
 CC produced as a result of synthetic mini-pro-insulin (B(1-30)-Arg-A(1-21))
 CC oligonucleotides being end ligated and inserted as 2 larger fragments IK1
 CC and IK2 into pUC19. pIK1 encodes B chain AAs 1-30 of human insulin
 CC When 2 inserts are re-isolated and ligated together into pUC19 plasmid
 CC pIK4 is produced via intermediate and ligated together into pUC19 plasmid
 CC (I) is useful in diabetes mellitus treatment it can be converted to human
 CC insulin directly or via intermediate mono-Arg-insulin (II).
 XX
 SQ Sequence 30 AA;
 Query Match 58.2%; Score 99; DB 11; Length 30;
 Best Local Similarity 69.0%; Pred. No. 6.1e-07;
 Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 VNFHLYGSHIREWLYLVAGERGFPDPT 30
 DB 2 VNFHLYGSHIREWLYLVAGERGFPDPT 30
 RESULT 20
 AAR30650
 ID AAR30650 standard; protein; 30 AA.
 AC AAR30650;
 XX
 DT 20-MAY-1998 (first entry)
 XX
 DE Phosphorylated human insulin B-chain.
 XX
 KW Diabetes mellitus; N-phosphorylated amino acid; blood glucose;
 KW control; insulin-dependent diabetes; isoelectric point.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers

modified_site 1 /note= "opt. phosphorylated"
 modified_site 3 /note= "opt. phosphorylated"
 modified_site 9 /note= "opt. phosphorylated"
 modified_site 16 /note= "opt. phosphorylated"
 modified_site 22 /note= "opt. phosphorylated"
 modified_site 26 /note= "opt. phosphorylated"
 modified_site 27 /note= "opt. phosphorylated"
 modified_site 29 /note= "opt. phosphorylated"
 modified_site 30 /note= "opt. phosphorylated"
 W09214754-A.
 03-SEP-1992.
 26-FEB-1992; 92WO-CA000082.
 26-FEB-1991; 91GB-0004037.
 (HOSP-) HOSPITAL SICK CHILDREN LOYAL TRUE BLUE.
 Loughheed WD;
 WPI; 1992-316123/38.
 phosphorylation of peptide(s) with phosphorus oxychloride - esp.
 for prodn. of phosphorylated insulin for diabetes treatment
 Claim 14-17; Page 10; 46pp; English.
 Phosphorylated insulin is claimed for treatment of diabetes in
 which at least one of the tyrosine residues is phosphorylated. The
 other residues indicated in the features table can also be
 phosphorylated. Phosphorylation is effected by reacting an aqueous
 solution of insulin with phosphorus oxychloride. The phosphorylated
 insulin is purified and only those portions which have
 substantially reduced isoelectric points are isolated. Unmodified
 insulin is removed to prevent negation of the benefit obtained from
 the "gentler pharmacokinetic properties" of the phosphorylated
 insulin.
 See also AAR30650 for phosphorylated A-chain.
 Sequence 30 AA;
 Query Match 58.2%; Score 99; DB 13; Length 30;
 Best Local Similarity 69.0%; Pred. No. 6.1e-07;
 Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 VNFHLYGSHIREWLYLVAGERGFPDPT 30
 DB 2 VNFHLYGSHIREWLYLVAGERGFPDPT 30
 RESULT 21
 AAR26470
 ID AAR26470 standard; Protein; 30 AA.
 AC AAR26470;
 XX
 DT 28-JAN-1993 (first entry)
 DE Sequence of human insulin B-chain.
 KW Serine protease; proteolytic enzyme; plastein reaction; insulin;
 KW SP 446 protease.

```
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FT Disulfide-bond 7 /note= "bonded to AA-7 of A-chain"
XX FT Disulfide-bond 19 /note= "bonded to AA-20 of A-chain"
XX FT Cleavage-site 13..14 /note= "see CC"
XX FT Cleavage-site 21..22 /note= "see CC"
XX PN W09213964-A.
XX PD 20-AUG-1992.
XX PF 06-FEB-1992; 92WO-DK00036.
XX PR 06-FEB-1991; 91DK-0000199.
XX PA (NOVO ) NOVO-NORDISK AS.
XX PI Budtz P, Nielsen P;
XX DR WPI; 1992-300049/36.
XX PT Increasing viscosity of protein prep(s), - by incubating with
XX PT serine protease specific for glutamic acid and aspartic acid
XX PT residues
XX PS Example; Fig 3; 38pp; English.
XX CC The serine protease of the invention has the following
XX CC characteristics: (a) it is a serine protease specific for glutamic
XX CC acid and aspartic acid residues; (b) it has a specific activity of
XX CC >25 CPU/g of enzyme protein.; (c) it has an apparent mol. wt. of
XX CC 23,600; (d) it is inhibited by DEP but not by PMSF, and (e) it
XX CC exhibits 75% or more of its maximising activity in the pH range 6.5-
XX CC 10. In partic. it is AAR26467. It was previously characterized in
XX CC US Patent No. 4,266,031 as a contaminant of subtilisin A. SP 446
XX CC (AAR26468) protease is an example of the claimed enzyme. The Glu
XX CC specificity of the SP 446 protease was determined using human insulin.
XX CC The cleavage sites in the insulin molecule are shown in
XX CC AAR26469, AAR26470.
XX SQ Sequence 30 AA;
Query Match 58.2%; Score 99; DB 13; Length 30;
Best Local Similarity 69.0%; Pred. No. 6.1e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
Db 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
RESULT 22
AAR44454
ID AAR44454 standard; protein; 30 AA.
XX AC AAR44454;
XX DT 10-JUN-1994 (first entry)
XX DE Human insulin B-chain.
XX KW anti-Insulin antibody; detection; peroxidase-labelled; protein A;
XX KW solid-phase; quantitation.
XX OS Homo sapiens.
XX PN JP05297000-A.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FT Disulfide-bond 7 /note= "bonded to AA-7 of A-chain"
XX FT Disulfide-bond 19 /note= "bonded to AA-20 of A-chain"
XX FT Cleavage-site 13..14 /note= "see CC"
XX FT Cleavage-site 21..22 /note= "see CC"
XX PN W09213964-A.
XX PD 20-AUG-1992.
XX PF 06-FEB-1992; 92WO-DK00036.
XX PR 06-FEB-1991; 91DK-0000199.
XX PA (NOVO ) NOVO-NORDISK AS.
XX PI Budtz P, Nielsen P;
XX DR WPI; 1992-300049/36.
XX PT Increasing viscosity of protein prep(s), - by incubating with
XX PT serine protease specific for glutamic acid and aspartic acid
XX PT residues
XX PS Example; Fig 3; 38pp; English.
XX CC The serine protease of the invention has the following
XX CC characteristics: (a) it is a serine protease specific for glutamic
XX CC acid and aspartic acid residues; (b) it has a specific activity of
XX CC >25 CPU/g of enzyme protein.; (c) it has an apparent mol. wt. of
XX CC 23,600; (d) it is inhibited by DEP but not by PMSF, and (e) it
XX CC exhibits 75% or more of its maximising activity in the pH range 6.5-
XX CC 10. In partic. it is AAR26467. It was previously characterized in
XX CC US Patent No. 4,266,031 as a contaminant of subtilisin A. SP 446
XX CC (AAR26468) protease is an example of the claimed enzyme. The Glu
XX CC specificity of the SP 446 protease was determined using human insulin.
XX CC The cleavage sites in the insulin molecule are shown in
XX CC AAR26469, AAR26470.
XX SQ Sequence 30 AA;
Query Match 58.2%; Score 99; DB 13; Length 30;
Best Local Similarity 69.0%; Pred. No. 6.1e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
Db 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
RESULT 22
AAR44454
ID AAR44454 standard; protein; 30 AA.
XX AC AAR44454;
XX DT 10-JUN-1994 (first entry)
XX DE Human insulin B-chain.
XX KW anti-Insulin antibody; detection; peroxidase-labelled; protein A;
XX KW solid-phase; quantitation.
XX OS Homo sapiens.
XX PN JP05297000-A.
```

```
XX 12-NOV-1993.
XX PD 17-APR-1992; 92JP-0122867.
XX PF 17-APR-1992; 92JP-0122867.
XX PR 17-APR-1992; 92JP-0122867.
XX PA (SRLS-) SRL KK.
XX DR WPI; 1993-397733/50.
XX XX Insulin antibody determ. in human body fluid sample - by
XX PT reacting sample and standard sample contg. non-human insulin
XX PT antibodies with insulin supported on a solid phase, and reacting
XX PT bound antibody with enzyme labelled protein
XX XX Claim 5; Fig 1; 9pp; Japanese.
XX PS Insulin (esp. human insulin) is attached to a solid phase for
XX CC reaction with a sample containing anti-insulin antibody. Bound
XX CC antibody is then reacted with enzyme-labelled protein A. The method
XX CC allows the concentration of anti-insulin antibody in a sample to be
XX CC determined. See also AAR44453 for the human insulin A-chain.
XX SQ Sequence 30 AA;
Query Match 58.2%; Score 99; DB 14; Length 30;
Best Local Similarity 69.0%; Pred. No. 6.1e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
Db 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
RESULT 23
AAR62642
ID AAR62642 standard; peptide; 30 AA.
XX AC AAR62642;
XX DT 05-JUL-1995 (first entry)
XX DE Bovine and human insulin B chain.
XX KW Porcine insulin B chain; insulin dependent diabetes; gamma interferon;
XX KW autoimmune disorders; pancreatic insulin producing cells; vaccine.
XX OS Bos taurus; Homo sapiens.
XX PN W09423737-A.
XX PD 27-OCT-1994.
XX PF 15-APR-1994; 94WO-US04179.
XX PR 16-APR-1993; 93US-0048979.
XX PA (RESE ) RESEARCH CORP TECHNOLOGIES INC.
XX DT Maclaren NK, Muir A;
XX DR WPI; 1994-341479/42.
XX PT Prevention of treatment of autoimmune diseases, esp. diabetes
XX PT using vaccine contg. antigen associated with disease-targeted
XX PT cells or tissues
XX PS Claim 6; Page 47; 65pp; English.
XX XX AAR62640-R62643 are insulin fragments, these can be combined with
XX CC Freund's incomplete adjuvant to produce an autoimmune disorder
XX CC vaccine. This vaccine is administered non-orally to pancreatic
```


Db 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

Search completed: December 23, 2002, 07:17:23
Job time : 37.4706 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:22 ; Search time 14.1176 Seconds
(without alignments)
62.524 Million cell updates/sec

Title: US-09-574-443-7_COPY_22_51
Perfect score: 170
Sequence: 1 DVNPHLGSHIREWLYLVAGERFNPDKT 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	58.2	30	1 US-08-212-696-2	Sequence 2, Appli
2	99	58.2	30	1 US-08-233-617-2	Sequence 2, Appli
3	99	58.2	30	1 US-08-160-376A-2	Sequence 2, Appli
4	99	58.2	30	1 US-08-301-838-2	Sequence 2, Appli
5	99	58.2	30	1 US-08-389-487-5	Sequence 5, Appli
6	99	58.2	30	2 US-08-508-664-7	Sequence 5, Appli
7	99	58.2	30	2 US-08-353-476-86	Sequence 7, Appli
8	99	58.2	30	2 US-08-353-476-86	Sequence 2, Appli
9	99	58.2	30	2 US-08-484-219-6	Sequence 2, Appli
10	99	58.2	30	2 US-08-992-676-2	Sequence 2, Appli
11	99	58.2	30	2 US-08-134-836-2	Sequence 4, Appli
12	99	58.2	30	3 US-08-750-391-4	Sequence 2, Appli
13	99	58.2	30	4 US-09-255-668-1	Sequence 1, Appli
14	99	58.2	30	4 US-09-099-307-2	Sequence 2, Appli
15	99	58.2	30	4 US-08-900-574-2	Sequence 2, Appli
16	99	58.2	30	4 US-09-386-303A-2	Sequence 2, Appli
17	99	58.2	30	4 US-09-201-227A-20	Sequence 20, Appli
18	99	58.2	30	5 PCT-US94-04179-3	Sequence 3, Appli
19	99	58.2	30	6 5169865-5	Patent No. 5169865
20	99	58.2	31	1 US-08-389-487-10	Sequence 10, Appli
21	99	58.2	31	2 US-08-992-676-5	Sequence 5, Appli
22	99	58.2	32	1 US-08-389-487-9	Sequence 9, Appli
23	99	58.2	32	2 US-08-992-676-4	Sequence 4, Appli
24	99	58.2	32	2 US-09-134-836-6	Sequence 6, Appli
25	99	58.2	32	4 US-09-386-303A-6	Sequence 6, Appli
26	99	58.2	36	1 US-08-030-731A-34	Sequence 34, Appli
27	99	58.2	51	1 US-07-707-542E-2	Sequence 2, Appli

Sequence 3, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 42, Appli
Sequence 9, Appli
Sequence 6, Appli
Sequence 8, Appli
Sequence 7, Appli
Sequence 11, Appli
Sequence 44, Appli
Sequence 6, Appli
Sequence 3, Appli
Sequence 5, Appli
Sequence 7, Appli
Sequence 1, Appli
Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-212-696-2
; Sequence 2, Application US/08212696
; Patent No. 5422339
; GENERAL INFORMATION:
; APPLICANT: George S. Eisenbarth et al.
; TITLE OF INVENTION: PEPTIDES HAVING INSULIN
; TITLE OF INVENTION: AUTOANTIBODY BUT NOT
; TITLE OF INVENTION: INSULIN RECEPTOR BINDING
; TITLE OF INVENTION: CAPACITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55Sx
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,696
; FILING DATE: 03/19/91
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/671,455
; FILING DATE: 03/19/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00303/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: linear
US-08-212-696-2

Query Match 58.2%; Score 99; DB 1; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

us-09-574-443-7_copy_22_51.dec23.ra1

Mon Dec 23 08:51:05 2002

TITLE OF INVENTION: Possessing Correctly Linked

TITLE OF INVENTION: Cystine Bridges

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenneth A. Genoni, Esq.

STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500

CITY: Somerville

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 08876-1258

COMPUTER READABLE FORM: DISKETTE, 3.5 INCH, 1.44 MB STORAGE

MEDIUM TYPE: IBM 386

OPERATING SYSTEM: WINDOWS 3.1

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/160,376A

FILING DATE: December 1, 1993

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GE P 4240420.7

FILING DATE: December 2, 1992

ATTORNEY/AGENT INFORMATION:

NAME: Barbara V. Maurer, Esq.

REGISTRATION NUMBER: 31,287

REFERENCE/DOCKET NUMBER: HOE 92/F 384

TELECOMMUNICATION INFORMATION:

TELEPHONE: (908) 231-4079

TELEFAX: (908) 231-2255

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 30 Amino Acids

TYPE: Amino Acid

TOPOLOGY: not relevant

US-08-160-376A-2

Query Match 58.2%; Score 99; DB 1; Length 30;

Best Local Similarity 69.0%; Pred. No. 4.7e-08;

Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30

DB 2 VNOHLCGSHLVLEALYLVCGRGFFFTPKT 30

RESULT 4

US-08-301-838-2

; Sequence 2, Application US/08301838

; Patent No. 5597796

; GENERAL INFORMATION:

; APPLICANT: Brange, Jens J.V.

; TITLE OF INVENTION: TRANSFERMAL INSULIN

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 55977960 No. 5597796disk of No. 5597796th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/301,838

; FILING DATE:

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/085,473

; FILING DATE: 30-JUN-1993

; PRIOR APPLICATION DATA:

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30

DB 2 VNOHLCGSHLVLEALYLVCGRGFFFTPKT 30

RESULT 2

US-08-233-617-2

; Sequence 2, Application US/08233617

; Patent No. 5466666

; GENERAL INFORMATION:

; APPLICANT: Obermeier, Rainer

; APPLICANT: Sabel, Walter

; APPLICANT: Dail, Peter

; APPLICANT: Geisen, Karl

; TITLE OF INVENTION: Amorphous Monospherical Forms of Insulin

; TITLE OF INVENTION: Derivatives

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; STREET: 1300 I Street, N.W., Suite 700

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/233,617

; FILING DATE: 25-APR-1994

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: P 43 13 702.4

; FILING DATE: 27-APR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Carol P. Elnaud

; REGISTRATION NUMBER: 32,220

; REFERENCE/DOCKET NUMBER: 02481.1374-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-408-4000

; TELEFAX: 202-408-4400

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 30 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORGANISM: Escherichia coli

; US-08-233-617-2

Query Match 58.2%; Score 99; DB 1; Length 30;

Best Local Similarity 69.0%; Pred. No. 4.7e-08;

Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30

DB 2 VNOHLCGSHLVLEALYLVCGRGFFFTPKT 30

RESULT 3

US-08-160-376A-2

; Sequence 2, Application US/08160376A

; Patent No. 5473049

; GENERAL INFORMATION:

; APPLICANT: Obermeier, Rainer

; APPLICANT: Gerl, Martin

; APPLICANT: Ludwig, Jürgen

; APPLICANT: Sabel, Walter

; TITLE OF INVENTION: Process For Obtaining Proinsulin

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:47 : Search time 82.3529 Seconds
(without alignments)
6.107 Million cell updates/sec

Title: US-09-574-443-7_COPY_22_51

Perfect score: 170

Sequence: 1 DVNFHLYGSHIREWLVLVAGERGFNDFPKT 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	58.2	30	9	US-09-947-563-2
2	99	58.2	30	10	US-09-853-844-2
3	99	58.2	30	10	US-09-815-229-2
4	99	58.2	30	10	US-09-815-229-15
5	99	58.2	32	9	US-09-947-563-6
6	99	58.2	32	10	US-09-815-229-14
7	99	58.2	51	9	US-10-028-410-3
8	99	58.2	52	9	US-10-054-873-5
9	99	58.2	56	10	US-09-815-229-10
10	99	58.2	86	9	US-09-878-380-1
11	99	58.2	86	9	US-10-028-410-2
12	99	58.2	86	9	US-10-054-873-4
13	99	58.2	96	9	US-09-947-563-5
14	99	58.2	96	9	US-09-947-563-5
15	99	58.2	107	9	US-10-054-873-6
16	99	58.2	110	9	US-09-804-409A-9
17	99	58.2	110	10	US-09-205-658-125
18	99	58.2	110	10	US-09-815-229-3
19	99	58.2	117	10	US-09-280-030-63

20	99	58.2	130	10	US-09-280-030-62
21	99	58.2	150	9	US-10-054-873-7
22	99	58.2	166	10	US-09-925-297-805
23	94	55.3	32	10	US-09-896-874-5
24	94	55.3	32	10	US-09-896-139-5
25	94	55.3	32	10	US-09-895-843-5
26	94	55.3	50	9	US-10-066-009A-3
27	94	55.3	124	10	US-09-894-711-18
28	86	50.6	124	10	US-09-736-611-12
29	86	50.6	124	10	US-09-740-359-12
30	86	50.6	124	10	US-09-894-711-12
31	86	50.6	125	10	US-09-736-611-10
32	86	50.6	125	10	US-09-740-359-10
33	86	50.6	125	10	US-09-894-711-10
34	86	50.6	144	10	US-09-736-611-6
35	86	50.6	144	10	US-09-740-359-5
36	86	50.6	144	10	US-09-894-711-7
37	86	50.6	146	10	US-09-894-711-5
38	86	50.6	147	10	US-09-736-611-8
39	86	50.6	147	10	US-09-740-359-7
40	83	48.8	50	10	US-09-853-844-4
41	76	44.7	51	10	US-09-853-844-3
42	70	41.2	46	10	US-09-205-658-137
43	68	40.0	46	10	US-09-205-658-132
44	68	40.0	46	10	US-09-205-658-133
45	65	38.2	46	10	US-09-205-658-134

ALIGNMENTS

RESULT 1
US-09-947-563-2
; Sequence 2, Application US/09947563
; Patent No. US20020156234A1

GENERAL INFORMATION:

APPLICANT: Rubroder, Franz-Josef
Keller, Reinhold

TITLE OF INVENTION: Improved process for obtaining

insulin precursors having correctly bonded cystine b

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESS: Finnegan, Henderson, Farrabow, Garrett &
Dunner

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/947,563

FILING DATE: 07-Sep-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/134,836

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Leslie McDonnell

REGISTRATION NUMBER: 34,872

REFERENCE/DOCKET NUMBER: 02481.1600-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 30 amino acids

TYPE: amino acid

STRANDEDNESS: single

;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; ORIGINAL SOURCE:
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: Protein
;; LOCATION: 1..30
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-947-563-2

Query Match 58.2%; Score 99; DB 9; Length 30;
Best Local Similarity 69.0%; Pred. No. 3.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || |||: | |||| |||||: |||
DB 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 2

US-09-853-844-2
; Sequence 2, Application US/09853844
; Patent No. US20020013269A1
; GENERAL INFORMATION:
; APPLICANT: Balschmidt, Per
; APPLICANT: Brange, Veilgaard Jens Jorgen
; TITLE OF INVENTION: Human Insulin Analogues
; FILE REFERENCE: 3343.270-US
; CURRENT APPLICATION NUMBER: US/09/853,844
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 08/965,221
; PRIOR FILING DATE: 1997-11-06
; PRIOR APPLICATION NUMBER: 08/531,842
; PRIOR FILING DATE: 1995-09-21
; PRIOR APPLICATION NUMBER: 08/275,196
; PRIOR FILING DATE: 1994-07-14
; PRIOR APPLICATION NUMBER: 07/976,805
; PRIOR FILING DATE: 1992-11-16
; PRIOR APPLICATION NUMBER: 07/453,445
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: 07/416,218
; PRIOR FILING DATE: 1989-10-02
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-853-844-2

Query Match 58.2%; Score 99; DB 10; Length 30;
Best Local Similarity 69.0%; Pred. No. 3.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || |||: | |||| |||||: |||
DB 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 3

US-09-815-229-2
; Sequence 2, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDERS
; FILE REFERENCE: P1786RIUS
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17

;; SEQ ID NO 2
;; LENGTH: 30
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-815-229-2

Query Match 58.2%; Score 99; DB 10; Length 30;
Best Local Similarity 69.0%; Pred. No. 3.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || |||: | |||| |||||: |||
DB 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 4

US-09-815-229-15
; Sequence 15, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDERS
; FILE REFERENCE: P1786RIUS
; CURRENT APPLICATION NUMBER: US/09/815,229
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 60/192,103
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 15
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Portion of insulin variant from EP 171.147 and EP 171.887.
US-09-815-229-15

Query Match 58.2%; Score 99; DB 10; Length 30;
Best Local Similarity 69.0%; Pred. No. 3.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || |||: | |||| |||||: |||
DB 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 5

US-09-947-563-6
; Sequence 6, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; TITLE OF INVENTION: insulin precursors having correctly bonded cystine
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/947,563
; FILING DATE: 07-Sep-2001
; CLASSIFICATION: <Unknown>

;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: 09/134,836
;; FILING DATE: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Leslie McDonnell
;; REGISTRATION NUMBER: 34,872
;; REFERENCE/DOCKET NUMBER: 02481.1600-00000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 408-4000
;; TELEFAX: (202) 408-4400
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 32 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; ORIGINAL SOURCE:
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: Protein
;; LOCATION: 1..32
;; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-947-563-6

Query Match 58.2%; Score 99; DB 9; Length 32;
Best Local Similarity 69.0%; Pred. No. 4e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGERGFNFDPKT 30
|| || |||: | |||| |||||: |||
Db 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 6
US-09-815-229-14
;; Sequence 14, Application US/09815229
;; Patent No. US20020058614A1
;; GENERAL INFORMATION:
;; APPLICANT: Filvaroff, Ellen H.
;; APPLICANT: Okumu, Franklin W.
;; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGENOUS DISORDERS
;; FILE REFERENCE: P1786RIUS
;; CURRENT APPLICATION NUMBER: US/09/815,229
;; CURRENT FILING DATE: 2001-03-22
;; PRIOR APPLICATION NUMBER: US 60/192,103
;; PRIOR FILING DATE: 2000-03-24
;; NUMBER OF SEQ ID NOS: 17
;; SEQ ID NO 14
;; LENGTH: 32
;; TYPE: PRT
;; ORGANISM: Artificial sequence
;; FEATURE:
;; OTHER INFORMATION: Portion of Insulin variant from EP 171,147.
US-09-815-229-14

Query Match 58.2%; Score 99; DB 10; Length 32;
Best Local Similarity 69.0%; Pred. No. 4e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGERGFNFDPKT 30
|| || |||: | |||| |||||: |||
Db 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 7
US-10-028-410-3
;; Sequence 3, Application US/10028410
;; Patent No. US20020160955A1
;; GENERAL INFORMATION:
;; APPLICANT: Dubaqui, Yves
;; APPLICANT: Lowman, Henry
;; TITLE OF INVENTION: PROTEIN VARIANTS

;; FILE REFERENCE: P1712R1-1
;; CURRENT APPLICATION NUMBER: US/10/028,410
;; CURRENT FILING DATE: 2001-12-19
;; PRIOR APPLICATION NUMBER: US/09/477,924
;; PRIOR FILING DATE: 2000-01-05
;; NUMBER OF SEQ ID NOS: 6
;; SEQ ID NO 3
;; LENGTH: 51
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-028-410-3

Query Match 58.2%; Score 99; DB 9; Length 51;
Best Local Similarity 69.0%; Pred. No. 6.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGERGFNFDPKT 30
|| || |||: | |||| |||||: |||
Db 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 8
US-10-054-873-5
;; Sequence 5, Application US/10054873
;; Patent No. US20020164712A1
;; GENERAL INFORMATION:
;; APPLICANT: Gan, Zhong Ru
;; TITLE OF INVENTION: Chimeric Protein Containing an Intramolecular Chaperone-Like Sequence
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Townsend and Townsend and Crew LLP
;; STREET: Two Embarcadero Center, Eighth Floor
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94111-3834
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/10/054,873
;; FILING DATE: 22-Jan-2002
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/CN98/00052
;; FILING DATE: 31-MAR-1998
;; APPLICATION NUMBER: US 09/423,100
;; FILING DATE: 11-DEC-2000
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Mycroft, Frank J
;; REGISTRATION NUMBER: 46,946
;; REFERENCE/DOCKET NUMBER: 020167-000130US
;; INFORMATION FOR SEQ ID NO: 5:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 52 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: <Unknown>
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-054-873-5

Query Match 58.2%; Score 99; DB 9; Length 52;
Best Local Similarity 69.0%; Pred. No. 6.8e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGERGFNFDPKT 30
|| || |||: | |||| |||||: |||
Db 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 9
US-09-815-229-10
; Sequence 10, Application US/09815229
; Patent No. US20020058614A1
; GENERAL INFORMATION:
; APPLICANT: Filvaroff, Ellen H.
; APPLICANT: Okumu, Franklin W.
; TITLE OF INVENTION: USE OF INSULIN FOR THE TREATMENT OF CARTILAGINOUS DISORDERS
; FILE REFERENCE: P1786RIUS
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US/09/815,229
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 10
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Portion of insulin variant from EP 171,886.
US-09-815-229-10

Query Match 58.2%; Score 99; DB 10; Length 56;
Best Local Similarity 69.0%; Pred. No. 7.4e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
DB 2 VNOHLCGSHLVEALYLVCGRGFFTYPTKT 30

RESULT 10
US-09-878-380-1
; Sequence 1, Application US/09878380
; Patent No. US20020160435A1
; GENERAL INFORMATION:
; APPLICANT: Fujirebio Inc.
; APPLICANT: KITAJIMA, Sachiko
; APPLICANT: KURANO, Yoshihiro
; APPLICANT: NAKATSUBO, Kaoru
; APPLICANT: NISHIZONO, Isao
; TITLE OF INVENTION: Immunoassay For Measuring Human C-Peptide and Kit Therefor
; FILE REFERENCE: 0760-0291P
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: JP 2000-174691
; PRIOR FILING DATE: 2000-06-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-878-380-1

Query Match 58.2%; Score 99; DB 9; Length 86;
Best Local Similarity 69.0%; Pred. No. 1.2e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
DB 2 VNOHLCGSHLVEALYLVCGRGFFTYPTKT 30

RESULT 11
US-10-028-410-2
; Sequence 2, Application US/10028410
; Patent No. US20020160955A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquié, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS

; FILE REFERENCE: P1712R1-1
; CURRENT APPLICATION NUMBER: US/10/028,410
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US/09/477,924
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 2
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-410-2

Query Match 58.2%; Score 99; DB 9; Length 86;
Best Local Similarity 69.0%; Pred. No. 1.2e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
DB 2 VNOHLCGSHLVEALYLVCGRGFFTYPTKT 30

RESULT 12
US-10-054-873-4
; Sequence 4, Application US/10054873
; Patent No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 86 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Query Match 58.2%; Score 99; DB 9; Length 86;
Best Local Similarity 69.0%; Pred. No. 1.2e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
DB 2 VNOHLCGSHLVEALYLVCGRGFFTYPTKT 30

RESULT 13
US-09-947-563-4
; Sequence 4, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rudroder, Franz-Josef
; Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; insulin precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrabow, Garrett &
; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/947,563
; FILING DATE: 07-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/134,836
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..96
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-947-563-4
Query Match 58.2%; Score 99; DB 9; Length 96;
Best Local Similarity 69.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 VNFHLYGSHIREWLYLVAGERGFFNPKT 30
Db 12 VNQHLCGSHLVREALYLVCGERGFFYTPKT 40
RESULT 14
US-09-947-563-5
; Sequence 5, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rudroder, Franz-Josef
; Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; insulin precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farrabow, Garrett &
; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/947,563
; APPLICATION NUMBER: 09/134,836
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..96
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-947-563-5
Query Match 58.2%; Score 99; DB 9; Length 96;
Best Local Similarity 69.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 VNFHLYGSHIREWLYLVAGERGFFNPKT 30
Db 12 VNQHLCGSHLVREALYLVCGERGFFYTPKT 40
RESULT 15
US-10-054-873-6
; Sequence 6, Application US/10054873
; Patent No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873

Query Match 58.2%; Score 99; DB 10; Length 110;
Best Local Similarity 69.0%; Pred. No. 1.6e-07;
Matches 20: Conservative 2; Mismatches 7; Indels

```
;
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Designated is
; OTHER INFORMATION: an amino acid sequence of
; OTHER INFORMATION: MWpmp10-Met-Proinsulin
US-09-280-030-63

Query Match          58.2%; Score 99; DB 10; Length 117;
Best Local Similarity 69.0%; Pred. No. 1.7e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
   |||||:| ||||| ||||| : |||
Db 33 VNQHLCGSHLVEALYLVCGERGFFYTPKT 61

RESULT 20
US-09-280-030-62
; Sequence 62, Application US/09280030A
; Patent No. US20010021515A1
; GENERAL INFORMATION:
; APPLICANT: Sato, Seiji
; APPLICANT: Higashikuni, Naohiko
; APPLICANT: Kudo, Toshiyuki
; APPLICANT: Kondo, Masaaki
; TITLE OF INVENTION: DNAS ENCODING NEW FUSION PROTEINS AND PROCESSES FOR
; TITLE OF INVENTION: PREPARING USEFUL POLYPEPTIDES THROUGH EXPRESSION OF THE
; TITLE OF INVENTION: DNAS
; FILE REFERENCE: 382.1026
; CURRENT APPLICATION NUMBER: US/09/280.030A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: JP10-87339/1998
; EARLIER FILING DATE: 1998-03-31
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Designated is
; OTHER INFORMATION: an amino acid sequence of
; OTHER INFORMATION: MWpmp10-(His)6-Linker-Met-Proinsulin
US-09-280-030-62

Query Match          58.2%; Score 99; DB 10; Length 130;
Best Local Similarity 69.0%; Pred. No. 1.9e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
   |||||:| ||||| ||||| : |||
Db 46 VNQHLCGSHLVEALYLVCGERGFFYTPKT 74

RESULT 21
US-10-054-873-7
; Sequence 7, Application US/10054873
; Patent No. US20020164712A1
; GENERAL INFORMATION:
; APPLICANT: Gan, Zhong Ru
; TITLE OF INVENTION: Chimeric Protein Containing an
; TITLE OF INVENTION: Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
;
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,873
; FILING DATE: 22-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO.PCT/CN98/00052
; FILING DATE: 31-MAR-1998
; APPLICATION NUMBER: US 09/423,100
; FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
; NAME: Mycroft, Frank J
; REGISTRATION NUMBER: 46,946
; REFERENCE/DOCKET NUMBER: 020167-0001300S
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-054-873-7

Query Match          58.2%; Score 99; DB 9; Length 150;
Best Local Similarity 69.0%; Pred. No. 2.2e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
   |||||:| ||||| ||||| : |||
Db 100 VNQHLCGSHLVEALYLVCGERGFFYTPKT 128

RESULT 22
US-09-925-297-805
; Sequence 805, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 805
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (84)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (105)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (132)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (144)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
```

; LOCATION: (145)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (165)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-925-297-805

Query Match 58.2%; Score 99; DB 10; Length 166;
Best Local Similarity 69.0%; Pred. No. 2.5e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 30
|| || |||: | |||| |||||: ||
Db 43 VNOHLCGSHLVEALYLVCGERGFFYTPK 71

RESULT 23
US-09-896-874-5
; Sequence 5, Application US/09896874
; Patent No. US20020016320A1
; GENERAL INFORMATION:
; APPLICANT: Fang, Lawrence Y.
; APPLICANT: John, Varghese
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE
; FILE REFERENCE: 13615.40USU1
; CURRENT APPLICATION NUMBER: US/09/896,874
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,323
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-896-874-5

Query Match 55.3%; Score 94; DB 10; Length 32;
Best Local Similarity 67.9%; Pred. No. 2.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
|| || |||: | |||| |||||: ||
Db 2 VNOHLCGSHLVEALYLVCGERGFFYTPK 29

RESULT 24
US-09-896-139-5
; Sequence 5, Application US/09896139
; Patent No. US2002012825A1
; GENERAL INFORMATION:
; APPLICANT: Beck, James P.
; APPLICANT: Fang, Lawrence Y.
; APPLICANT: Freskos, John N.
; APPLICANT: Gailunas, Andrea
; APPLICANT: Hom, Roy
; APPLICANT: Jagodzinska, Barbara
; APPLICANT: John, Varghese
; APPLICANT: Mailalard, Michel
; APPLICANT: Pulley, Shon R.
; APPLICANT: Tenbrink, Ruth E.
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE
; FILE REFERENCE: 13615.25USU4
; CURRENT APPLICATION NUMBER: US/09/896,139
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,323
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 60/252,736
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/255,956
; PRIOR FILING DATE: 2000-12-15

; PRIOR APPLICATION NUMBER: US 60/268,497
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/279,779
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/295,589
; PRIOR FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-896-139-5

Query Match 55.3%; Score 94; DB 10; Length 32;
Best Local Similarity 67.9%; Pred. No. 2.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
|| || |||: | |||| |||||: ||
Db 2 VNOHLCGSHLVEALYLVCGERGFFYTPK 29

RESULT 25
US-09-895-843-5
; Sequence 5, Application US/09895843
; Patent No. US20020143177A1
; GENERAL INFORMATION:
; APPLICANT: Beck, James P.
; APPLICANT: Fang, Lawrence Y.
; APPLICANT: Freskos, John N.
; APPLICANT: Gailunas, Andrea
; APPLICANT: Hom, Roy
; APPLICANT: Jagodzinska, Barbara
; APPLICANT: John, Varghese
; APPLICANT: Mailalard, Michel
; APPLICANT: Pulley, Shon R.
; APPLICANT: Tenbrink, Ruth E.
; TITLE OF INVENTION: COMPOUNDS TO TREAT ALZHEIMER'S DISEASE
; FILE REFERENCE: 13615.41USU1
; CURRENT APPLICATION NUMBER: US/09/895,843
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,323
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-895-843-5

Query Match 55.3%; Score 94; DB 10; Length 32;
Best Local Similarity 67.9%; Pred. No. 2.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
|| || |||: | |||| |||||: ||
Db 2 VNOHLCGSHLVEALYLVCGERGFFYTPK 29

Search completed: December 23, 2002, 07:20:19
Job time : 83.3529 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:17 ; Search time 15.2941 Seconds
(without alignments)
188.571 Million cell updates/sec

Title: US-09-574-443-7_COPY_22_51
Perfect score: 170
Sequence: 1 DVNFHLYGSHIREWLYLVAGERGFNDPKT 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	58.2	51	1 INEL	insulin - elephant
2	99	58.2	110	1 IPHU	insulin precursor
3	99	58.2	110	2 B42179	insulin precursor
4	99	58.2	110	2 A42179	insulin precursor
5	99	58.2	110	2 JQ0178	insulin precursor
6	96	56.5	51	1 INGS	insulin - goose
7	96	56.5	81	1 IPDK	insulin precursor
8	95	55.9	51	1 INHY	insulin - hamster
9	95	55.9	103	2 IS1221	insulin precursor
10	95	55.9	110	1 INRB	insulin precursor
11	95	55.9	110	2 I48166	insulin precursor
12	94	55.3	51	1 INWHP	insulin - sperm wh
13	94	55.3	51	1 INWHF	insulin - finback
14	94	55.3	51	1 INWH1S	insulin - sei whal
15	94	55.3	51	1 INGT	insulin - goat
16	94	55.3	51	1 INCT	insulin - cat
17	94	55.3	51	2 JQ0362	insulin - North Am
18	94	55.3	51	2 A59151	insulin precursor
19	94	55.3	77	1 INSH	insulin precursor
20	94	55.3	84	1 IPGC	insulin precursor
21	94	55.3	86	1 IPHO	insulin precursor
22	94	55.3	96	2 PC7082	epidermal growth f
23	94	55.3	105	1 IPBO	insulin precursor
24	94	55.3	110	1 IPDG	insulin precursor
25	93	54.7	51	1 INKSKQ	insulin - common s
26	93	54.7	108	2 A39883	insulin precursor
27	92	54.1	51	1 INMSSP	insulin - Egyptian
28	91	53.5	51	1 INTK	insulin - turkey (
29	91	53.5	51	1 INOS	insulin - ostrich

30 91 53.5 51 1 A61129
31 91 53.5 51 2 A60414
32 91 53.5 107 1 IPCH
33 90 52.9 50 1 INEN2
34 90 52.9 51 1 INCM2
35 89 52.4 106 1 IPXL1
36 85 50.0 52 2 S44469
37 85 50.0 52 2 S44470
38 85 50.0 106 1 IPXL2
39 84 49.4 52 1 INXGA
40 84 49.4 108 1 INMS1
41 84 49.4 110 1 IPT1
42 83 48.8 50 1 INFIS
43 83 48.8 50 1 INON
44 83 48.8 50 1 INONC
45 83 48.8 51 1 INCB

insulin - black-be
insulin - slider t
insulin precursor
insulin - skipjack
insulin - Arabian
insulin I precursor
insulin I1 - North
insulin I2 - North
insulin II precurs
insulin - alligato
insulin I precursor
insulin I precursor
insulin - shorthor
insulin [validated
insulin - coho sal
insulin - Chinchil

ALIGNMENTS

RESULT 1

INEL

insulin - elephant

C:Species: Elephantidae gen. sp. (elephant)

C:Date: 24-Apr-1984 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999

C:Accession: A01584

R:Smith, L.F.

Am. J. Med. 40, 662-666, 1966

A:Title: Species variation in the amino acid sequence of insulin.

A:Reference number: A90029; MUID:66160119; PMID:5949593

A:Accession: A01584

A:Molecule type: protein

A:Residues: 1-30:31-51 <SMI>

A:Note: the species of elephant is not given, but it is most probably the Indian

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:1-30,31-51/Product: insulin #status experimental <MAT>

F:31-51/Domain: insulin chain A #status experimental <ACH>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 58.2%; Score 99; DB 1; Length 51;

Best Local Similarity 69.0%; Pred. No. 3e-07;

Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 2 VNFHLYGSHIREWLYLVAGERGFNDPKT 30

|| || |||: | ||| ||||: |||

Db 2 VNOHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 2

IPHU

insulin precursor [validated] - human

N:Alternate names: preproinsulin

C:Species: Homo sapiens (man)

C:Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 08-Dec-2000

C:Accession: A93222; A93216; A93214; A92075; A91186; I58114; A015

R:Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.

Nature 284, 26-32, 1980

A:Title: Sequence of the human insulin gene.

A:Reference number: A93222; MUID:80120725; PMID:6243748

A:Accession: A93222

A:Molecule type: DNA

A:Residues: 1-110 <BEL>

A:Cross-references: GB:J00265; NID:9186429; PIDN:AA59172.1; PID:9386828

R:Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.

Science 209, 612-615, 1980

A:Title: Genetic variation in the human insulin gene.

A:Reference number: A94253; MUID:80236313; PMID:6248962

A:Accession: A94253

A:Molecule type: DNA

A:Residues: 1-110 <ULL>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828
R;Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.
Nature 282, 525-527, 1979
A;Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.
A;Reference number: A93216; MUID:80054779; PMID:503234
A;Accession: A93216
A;Molecule type: mRNA
A;Residues: 1-110 <BEL2>
A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828
R;Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.
Science 208, 57-59, 1980
A;Title: Nucleotide sequence of human preproinsulin complementary DNA.
A;Reference number: A94251; MUID:80147417; PMID:6927840
A;Accession: A94251
A;Molecule type: mRNA
A;Residues: 1-110 <SUR>
A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828
R;Nicoll, D.S.H.W.; Smith, L.F.
Nature 187, 483-485, 1960
A;Title: Amino acid sequence of human insulin.
A;Reference number: A93144
A;Accession: A93144
A;Molecule type: protein
A;Residues: 25-54;90-110 <NIC>
R;Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.
J. Biol. Chem. 246, 1375-1386, 1971
A;Title: Studies on human proinsulin. Isolation and amino acid sequence of the human part
A;Reference number: A92075; MUID:71116410; PMID:5101771
A;Accession: A92075
A;Molecule type: protein
A;Residues: 57-87 <OYE>
R;Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
Eur. J. Biochem. 20, 190-199, 1971
A;Title: Amino acid sequence of the C-peptide of human proinsulin.
A;Reference number: A91186; MUID:71257722; PMID:5560404
A;Accession: A91186
A;Molecule type: protein
A;Residues: 57-87 <ROA>
R;Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boltard, C.; Froguel, P.; Lathrop, M.; Bell
Nature Genet. 4, 305-310, 1993
A;Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment
A;Reference number: I58114; MUID:93364428; PMID:8358440
A;Accession: I58114
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-59,63-110 <RES>
A;Cross-references: GB:LL5440; NID:g307071; PIDN:AAA59179.1; PID:g307072
R;Stieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.
Helv. Chim. Acta 57, 2617-2621, 1974
A;Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.
A;Reference number: A91636; MUID:75077277; PMID:4443293
A;Contents: annotation; synthesis
A;Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical
A;Note: article in German with English abstract
R;Naithani, V.K.
Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
A;Title: The synthesis of C-peptide of human proinsulin.
A;Reference number: A91658; MUID:75040007; PMID:4803504
A;Contents: annotation; synthesis of residues 57-87
R;Geiger, R.; Jaeger, G.; Koenig, W.
Chem. Ber. 106, 2347-2352, 1973
A;Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9
A;Reference number: A90914
A;Contents: annotation; synthesis of residues 57-87
R;Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.
Biochem. J. 310, 869-874, 1995
A;Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide junction
A;Reference number: S58661; MUID:96013185; PMID:7575420
A;Contents: annotation; site-directed mutagenesis study of proteolytic processing
C;Genetics:
A;Gene: GDB:INS
A;Cross-references: GDB:119349; OMIM:176730
A;Map position: 11p15.5-11p15.5

A;Introns: 63/1
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-54/Domain: insulin chain B #status experimental <BCH>
F;25-54,90-110/Product: insulin #status experimental <MAT>
F;57-87/Domain: connecting C peptide #status experimental <CPEP>
F;90-110/Domain: insulin chain A #status experimental <ACH>
F;31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 58.28; Score 99; DB 1; Length 110;
Best Local Similarity 69.08; Pred. No. 6.9e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
II IIII: I IIII IIII: III
DB 26 VNHLCGSHLVEALYLCGERGFFYTPKT 54

RESULT 3

B42179

insulin precursor - green monkey

C;Species: Cercopithecus aethiops (green monkey, grivet)

C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999

C;Accession: B42179; A05232; S16494; S22056

R;Seino, S.; Bell, G.I.; Li, W.H.

Mol. Biol. Evol. 9, 193-203, 1992

A;Title: Sequences of primate insulin genes support the hypothesis of a slower rat

A;Reference number: A42179; MUID:92219953; PMID:1560757

A;Accession: B42179

A;Molecule type: DNA

A;Residues: 1-110 <SEI>

A;Cross-references: EMBL:X61092; NID:g22808; PIDN:CAA43405.1; PID:g22809

A;Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBIP:95194)

R;Peterson, J.D.; Neuhlich, S.; Oyer, P.E.; Steiner, D.F.

J. Biol. Chem. 247, 4866-4871, 1972

A;Title: Determination of the amino acid sequence of the monkey, sheep, and dog p

A;Reference number: A92111; MUID:72258016; PMID:4626369

A;Accession: A05232

A;Molecule type: protein

A;Residues: 57-87 <PET>

C;Genetics:

A;Introns: 63/1

C;Superfamily: insulin

C;Keywords: hormone; pancreas

F;1-24/Domain: signal sequence #status predicted <SIG>

F;25-54/Domain: insulin chain B #status predicted <BCH>

F;25-54,90-110/Product: insulin #status predicted <MAT>

F;57-87/Domain: connecting peptide #status experimental <CPEP>

F;90-110/Domain: insulin chain A #status predicted <ACH>

F;31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 58.28; Score 99; DB 2; Length 110;

Best Local Similarity 69.08; Pred. No. 6.9e-07;

Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
II IIII: I IIII IIII: III
DB 26 VNHLCGSHLVEALYLCGERGFFYTPKT 54

RESULT 4

A42179

insulin precursor - chimpanzee

C;Species: Pan troglodytes (chimpanzee)

C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999

C;Accession: A42179; S22058

R;Seino, S.; Bell, G.I.; Li, W.H.

Mol. Biol. Evol. 9, 193-203, 1992

A;Title: Sequences of primate insulin genes support the hypothesis of a slower rat

A;Reference number: A42179; MUID:92219953; PMID:1560757

A;Accession: A42179

A;Status: preliminary

A:Molecule type: DNA
A:Residues: 1-110 <SEI>
A:Cross-references: EMBL:X61089; NID:g38251; PIDN:CAA43403.1; PID:g38252
A:Note: sequence extracted from NCBI backbone (NCBIP:95067)
C:Genetics:
A:introns: 63/1
C:Superfamily: insulin

Query Match 58.2%; Score 99; DB 2; Length 110;
Best Local Similarity 69.0%; Pred. No. 6.9e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
Qy 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
Db 26 VQHLGSHLVEALYLVCGERGFFYTPKT 54

RESULT 5
JQ0178
insulin precursor - crab-eating macaque
C:Species: Macaca fascicularis (crab-eating macaque)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 16-Jul-1999
C:Accession: JQ0178
R:Wetckam, W.; Gronenberg, J.; Lelneweber, M.; Wengenmayer, F.; Winnacker, E.L.
Gene 19, 179-183, 1982
A:Title: The nucleotide sequence of cDNA coding for preproinsulin from the primate Macaca
A:Reference number: JQ0178; MUID:83080474; PMID:6184262
A:Accession: JQ0178
A:Molecule type: mRNA
A:Residues: 1-110 <WET>
A:Cross-references: GB:J00336; NID:g342121; PIDN:AAA36849.1; PID:g342122
C:Superfamily: insulin
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54,90-110/Product: insulin #status predicted <MAT>
F:25-54/Domain: insulin chain B #status predicted <BCH>
F:55-89/Domain: insulin connecting C peptide #status predicted <CPT>
F:90-110/Domain: insulin chain A #status predicted <ACH>
F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 58.2%; Score 99; DB 2; Length 110;
Best Local Similarity 69.0%; Pred. No. 6.9e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
Qy 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
Db 26 VQHLGSHLVEALYLVCGERGFFYTPKT 54

RESULT 6
INGS
insulin - goose
C:Species: Anser anser (domestic goose)
C:Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C:Accession: J00007
R:Xu, Y.; Lin, N.; Zhang, Y.; Zhang, Y.
Rexue Tongbao 28, 966-968, 1983
A:Title: Isolation and sequence determination of goose insulin.
A:Reference number: J00007
A:Accession: J00007
A:Molecule type: protein
A:Residues: 1-30;31-51 <XUY>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 56.5%; Score 96; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 8.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 3 NPHLYGSHIREWLYLVAGRGFNFDPKT 30

Db 3 NQHLGSHLVEALYLVCGERGFFYSPKT 30
|| || ||| : |||| |||| : |||

RESULT 7
IPDK
insulin precursor - duck (tentative sequence)
C:Species: Anas platyrhynchos (domestic duck)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C:Accession: A01600; A91205
R:Markussen, J.; Sundby, F.
Int. J. Pept. Protein Res. 5, 37-48, 1973
A:Title: Duck insulin: isolation, crystallization and amino acid sequence.
A:Reference number: A91766; MUID:74055140; PMID:4763354
A:Accession: A01600
A:Molecule type: protein
A:Residues: 1-30;61-81 <MAR>
A:Experimental source: Pekin breed
R:Markussen, J.; Sundby, F.
Eur. J. Biochem. 34, 401-408, 1973
A:Title: Isolation and amino-acid sequence of the C-peptide of duck proinsulin.
A:Reference number: A91205; MUID:73210102; PMID:4715652
A:Accession: A91205
A:Molecule type: protein
A:Residues: 33-58 <MAR>
C:Comment: X's at positions 31-32 and 59-60 represent paired basic residues predi
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30,61-81/Product: insulin #status experimental <MAT>
F:33-58/Domain: connecting peptide #status experimental <CPEP>
F:61-81/Domain: insulin chain A #status experimental <ACH>
F:7-67,19-80,66-71/Disulfide bonds: #status predicted

Query Match 56.5%; Score 96; DB 1; Length 81;
Best Local Similarity 67.9%; Pred. No. 1.3e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
Qy 3 NPHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
Db 3 NQHLGSHLVEALYLVCGERGFFYSPKT 30

RESULT 8
INH
insulin - hamster
C:Species: Cricetinae gen. sp. (hamster)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A91456
R:Neelon, F.A.; Delcher, H.K.; Steinman, H.; Lebovitz, H.E.
Fed. Proc. 32, 300, 1973
A:Title: Structure of hamster insulin: comparison with a tumor insulin.
A:Reference number: A91456
A:Accession: A91456
A:Molecule type: protein
A:Residues: 1-30;31-51 <NEE>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 55.9%; Score 95; DB 1; Length 51;
Best Local Similarity 65.5%; Pred. No. 1.1e-06;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
Db 2 VQHLGSHLVEALYLVCGERGFFYTPKS 30

RESULT 9

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999

C; Keyw

C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999

F;1-30,31-51/Product: Insulin #status experimental <MAT>
F;31-51/Domain: Insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 55.3%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 1.6e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
DB 2 VNHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 14

INWH15

Insulin - sei whale

C:Species: Balaenoptera borealis (sei whale)

C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999

C:Accession: A01582

R:Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.

Nature 181, 1468-1469, 1958

A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin.

A:Reference number: A93142

A:Accession: A01582

A:Molecule type: protein

A:Residues: 1-30;31-51 <ISR>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F;1-30/Domain: Insulin chain B #status experimental <BCH>

F;1-30,31-51/Product: Insulin #status experimental <MAT>

F;31-51/Domain: Insulin chain A #status experimental <ACH>

F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 55.3%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 1.6e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
DB 2 VNHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 15

INCT

Insulin - goat

C:Species: Capra aegagrus hircus (domestic goat)

C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999

C:Accession: A01586

R:Smith, L.F.

Am. J. Med. 40, 662-666, 1966

A:Title: Species variation in the amino acid sequence of insulin.

A:Reference number: A90029; MUID:66160119; PMID:5949593

A:Accession: A01586

A:Molecule type: protein

A:Residues: 1-30;31-51 <SMI>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F;1-30/Domain: Insulin chain B #status experimental <BCH>

F;1-30,31-51/Product: Insulin #status experimental <MAT>

F;31-51/Domain: Insulin chain A #status experimental <ACH>

F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 55.3%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 1.6e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
DB 2 VNHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 16

INCT

Insulin - cat

C:Species: Felis silvestris catus (domestic cat)

C:Date: 30-Jun-1987 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999

C:Accession: A01588

R:Halliden, G.; Gafvelin, G.; Mutt, V.; Jornvall, H.

Arch. Biochem. Biophys. 247, 20-27, 1986

A:Title: Characterization of cat insulin.

A:Reference number: A01588; MUID:86214076; PMID:3518635

A:Accession: A01588

A:Molecule type: protein

A:Residues: 1-30;31-51 <HAL>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F;1-30/Domain: Insulin chain B #status experimental <BCH>

F;1-30,31-51/Product: Insulin #status experimental <MAT>

F;31-51/Domain: Insulin chain A #status experimental <ACH>

F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 55.3%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 1.6e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
DB 2 VNHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 17

JQ0362

Insulin - North American opossum

C:Species: Didelphis virginiana, Didelphis marsupialis virginiana (North American

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 23-Aug-1997

C:Accession: JQ0362; JQ0363

R:Yu, J.H.; Eng, J.; Rattan, S.; Yalow, R.S.

Peptides 10, 1195-1197, 1989

A:Title: Opossum insulin, glucagon and pancreatic polypeptide: amino acid sequence

A:Reference number: JQ0362; MUID:90160042; PMID:2695899

A:Accession: JQ0362

A:Molecule type: protein

A:Residues: 1-51 <YUJ>

C:Superfamily: Insulin

C:Keywords: hormone; pancreas

F;1-30/Domain: Insulin chain B #status experimental <BCI>

F;1-30,31-51/Product: Insulin #status experimental <MAT>

F;31-51/Domain: Insulin chain A #status experimental <ACI>

F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 55.3%; Score 94; DB 2; Length 51;
Best Local Similarity 67.9%; Pred. No. 1.6e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
DB 2 VNHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 18

A59151

Insulin precursor - jack bean (fragments)

N:Alternate names: hypoglycemic agent; plant insulin

C:Species: Canavalia ensiformis (jack bean)

C:Date: 07-Dec-1999 #sequence_revision 07-Dec-1999 #text_change 10-Dec-1999

C:Accession: B59151; A59151

R:Oliveira, A.E.A.; Machado, O.L.T.; Gomes, V.M.; Xavier-Neto, J.; Pereira, A.C.P

Protein Pept. Lett. 6, 15-21, 1999

A:Title: Jack bean seed coat contains a protein with complete sequence homology t

A:Reference number: A59151

A:Accession: B59151

A:Molecule type: protein

A:Residues: 1-30 <MACB>

A:Accession: A59151

A:Molecule type: protein

A:Residues: 31-51 <MACA>

C:Comment: The two chains are probably produced from the same precursor.

C:Superfamily: insulin

F:1-30,31-51/Product: insulin #status experimental <MAT>

F:1-30/Domain: chain B #status experimental <CHB>

F:31-51/Domain: chain A #status experimental <CHA>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 55.3%; Score 94; DB 2; Length 51;

Best Local Similarity 67.9%; Pred. No. 1.6e-06;

Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29

II III III: I IIII IIIII: II

Db 2 VNOHLCGSHLVREALYLVCGRGFFFTPK 29

RESULT 19

INSH

insulin precursor - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999

C:Accession: S16430; S16431

R:Brown, H.; Sanger, F.; Kitai, R.

Biochem. J. 60, 556-565, 1955

A:Title: The structure of pig and sheep insulins.

A:Reference number: A90344

A:Accession: S16430

A:Molecule type: protein

A:Residues: 1-30:57-77 <BRO>

R:Peterson, J.D.; Nehrlich, S.; Oyer, P.E.; Steiner, D.F.

J. Biol. Chem. 247, 4866-4871, 1972

A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin

A:Reference number: A92111; MUID:72258016; PMID:4626369

A:Accession: S16431

A:Molecule type: protein

A:Residues: 31-56 <PET>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:31-57/Product: insulin #status experimental <MAT>

F:57-77/Domain: connecting peptide #status experimental <CPEP>

F:7-63,19-76,62-67/Disulfide bonds: #status predicted

Query Match 55.3%; Score 94; DB 1; Length 77;

Best Local Similarity 67.9%; Pred. No. 2.4e-06;

Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29

II III III: I IIII IIIII: II

Db 2 VNOHLCGSHLVREALYLVCGRGFFFTPK 29

RESULT 20

IPPG

insulin precursor - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 22-Jun-1991 #sequence_revision 22-Jun-1991 #text_change 16-Jul-1999

C:Accession: A01583; A94572; S16492; A60835; B60835

R:Chance, R.E.; Ellis, R.M.; Bromer, W.W.

Science 161, 165-167, 1968

A:Title: Porcine proinsulin: characterization and amino acid sequence.

A:Reference number: A94240; MUID:68286485; PMID:5657063

A:Accession: A01583

A:Molecule type: protein

A:Residues: 1-34,'Q',36-84 <CHA>

R:Chance, R.E.

submitted to the Atlas, July 1970

A:Reference number: A94572

A:Accession: A94572

A:Molecule type: protein

A:Residues: 1-84 <CH2>

R:Brown, H.; Sanger, F.; Kitai, R.

Biochem. J. 60, 556-565, 1955

A:Title: The structure of pig and sheep insulins.

A:Reference number: A90344

A:Accession: S16492

A:Molecule type: protein

A:Residues: 1-30;31-51 <BRO>

R:Snell, L.; Damgaard, U.

Horm. Metab. Res. 20, 476-480, 1988

A:Title: Proinsulin heterogeneity in pigs.

A:Reference number: A60835; MUID:890321178; PMID:3181865

A:Accession: A60835

A:Molecule type: protein

A:Residues: 33-38,40-62 <SNE>

A:Note: the authors report the characterization of a connecting peptide variant

A:Accession: B60835

A:Molecule type: protein

A:Residues: 33-62 <SN2>

R:Blundell, T.; Dodson, G.; Hodgkin, D.; Mercola, D.

Adv. Protein Chem. 26, 279-402, 1972

A:Title: Insulin: the structure in the crystal and its reflection in chemistry and

A:Reference number: A90017

C:Contents: annotation; X-ray crystallography, 1.9 angstroms

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:31-63/Domain: connecting peptide #status experimental <MAT>

F:64-84/Domain: insulin chain A #status experimental <ACH>

F:7-70,19-83,69-74/Disulfide bonds: #status experimental

Query Match

Best Local Similarity 55.3%; Score 94; DB 1; Length 84;

Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29

II III III: I IIII IIIII: II

Db 2 VNOHLCGSHLVREALYLVCGRGFFFTPK 29

RESULT 21

IPHO

insulin precursor - horse

C:Species: Equus caballus (domestic horse)

C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999

C:Accession: A01580; A92120

R:Harris, J.I.; Sanger, F.; Naughton, M.A.

Arch. Biochem. Biophys. 65, 427-428, 1956

A:Title: Species differences in insulin.

A:Reference number: A90082

A:Accession: A01580

A:Molecule type: protein

A:Residues: 1-30:66-86 <HAR>

R:Tagger, H.S.; Steiner, D.F.

J. Biol. Chem. 247, 7936-7940, 1972

A:Title: Primary structures of the proinsulin connecting peptides of the rat and

A:Reference number: A92120; MUID:73061498; PMID:4640931

A:Accession: A92120

A:Molecule type: protein

A:Residues: 33-63 <TAG>

C:Comment: X's at positions 31-32 and 64-65 represent paired basic residues assume

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:31-63/Product: insulin #status experimental <MAT>

F:66-86/Domain: connecting peptide #status experimental <CPEP>

F:7-72,19-85,71-76/Disulfide bonds: #status predicted

Query Match

Best Local Similarity 55.3%; Score 94; DB 1; Length 86;

Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29

Db 2 VNOHLCGSHLVYALVLCGERGFFYTPK 29

RESULT 22

PC7082
Epidermal growth factor/single chain insulin fusion protein - Bacillus brevis (fragment)
C:Species: Bacillus brevis
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C:Accession: PC7082; PC7083
R:Koh, M.; Hanagata, H.; Ebisu, S.; Morihara, K.; Takagi, H.
Biosci. Biotechnol. Biochem. 64, 1079-1081, 2000
A:Title: Use of Bacillus brevis for synthesis and secretion of Des-B30 single-chain human insulin
A:Reference number: PC7082; MUID:20335834; PMID:10879487
A:Accession: PC7082
A:Molecule type: DNA
A:Residues: 1-96 <KOH>
A:Accession: PC7083
A:Molecule type: protein
A:Residues: 19-28 <K02>
C:Genetics:
C:Gene: egf-sci
C:Superfamily: Insulin
C:Keywords: fusion protein

Query Match 55.3%; Score 94; DB 2; Length 96;

Best Local Similarity 67.9%; Pred. No. 3.1e-06;

Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29

Db 48 VNOHLCGSHLVYALVLCGERGFFYTPK 75

RESULT 23

IPBO
Insulin precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1984 #sequence_revision 22-Apr-1995 #text_change 16-Jul-1999
C:Accession: A40909; A92080; A92074; A91185; A90342; A90341; S48185; S46258; A01587
R:D'Agostino, J.; Younes, M.A.; White, J.W.; Besch, P.K.; Field, J.B.; Frazier, M.L.
Mol. Endocrinol. 1, 327-331, 1987
A:Title: Cloning and nucleotide sequence analysis of complementary deoxyribonucleic acid (cDNA) for the insulin precursor gene in the bovine pancreas
A:Reference number: A40909; MUID:88288209; PMID:2456452
A:Accession: A40909
A:Molecule type: mRNA
A:Residues: 1-105 <DAA>
A:Cross-references: GB:M54979; NID:g163578; PIDN:AAA30722.1; PID:g163579
A:Experimental source: fetal pancreas
R:Nolan, C.; Margolliash, E.; Peterson, J.D.; Steiner, D.F.
J. Biol. Chem. 246, 2780-2795, 1971
A:Title: The structure of bovine proinsulin.
A:Reference number: A92080; MUID:71166442; PMID:4928892
A:Accession: A92080
A:Molecule type: protein
A:Residues: 25-105 <NOL>
R:Steiner, D.F.; Cho, S.; Over, P.E.; Terris, S.; Peterson, J.D.; Rubenstein, A.H.
J. Biol. Chem. 246, 1365-1374, 1971
A:Title: Isolation and characterization of proinsulin C-peptide from bovine pancreas.
A:Reference number: A92074; MUID:71116409; PMID:5545080
A:Accession: A92074
A:Molecule type: protein
A:Residues: 57-82 <STE>
R:Salokangas, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
Eur. J. Biochem. 20, 183-189, 1971
A:Title: Bovine proinsulin: amino acid sequence of the C-peptide isolated from pancreas.
A:Reference number: A91185; MUID:71257721; PMID:5105368
A:Accession: A91185
A:Molecule type: protein
A:Residues: 57-82 <SAL>
R:Sanger, F.; Thompson, E.O.P.
Biochem. J. 53, 366-374, 1953
A:Title: The amino-acid sequence in the glycol chain of insulin. 2. The investigation of

A:Reference number: A90342
A:Accession: A90342
A:Molecule type: protein
A:Residues: 85-105 <SAN>
R:Sanger, F.; Tuppy, H.
Biochem. J. 49, 481-490, 1951
A:Title: The amino-acid sequence in the phenylalanyl chain of insulin. 2. The inv

A:Reference number: A90341

A:Accession: A90341

A:Molecule type: protein

A:Residues: 25-54 <SA2>

R:Cheng, R.; Kawakishi, S.

Eur. J. Biochem. 223, 759-764, 1994

A:Title: Site-specific oxidation of histidine residues in glycosylated insulin mediat

A:Reference number: S48184; MUID:94333378; PMID:8055951

A:Accession: S48184

A:Molecule type: protein

A:Residues: 85-105 <CHE>

A:Accession: S48185

A:Status: preliminary

A:Molecule type: protein

A:Residues: 25-30, 'X', 32-42, 'X', 44-54 <CH2>

R:Ryle, A.P.; Sanger, F.; Smith, L.F.; Kitai, R.

Biochem. J. 60, 541-556, 1955

A:Title: The disulphide bonds of insulin.

A:Reference number: A90343

A:Contents: annotation; amides; disulfides

R:Wenzel, T.; Eckerskorn, C.; Lottspeich, F.; Baumeister, W.

FEBS Lett. 349, 205-209, 1994

A:Title: Existence of a molecular ruler in proteasomes suggested by analysis of d

A:Reference number: S46258; MUID:94326921; PMID:8050567

A:Accession: S46258

A:Status: preliminary

A:Molecule type: protein

A:Residues: 25-54 <WEN>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-54/Domain: insulin chain B #status experimental <BCH>

F:57-82/Domain: connecting peptide #status experimental <MAT>

F:85-105/Domain: insulin chain A #status experimental <CPEP>

F:91-91,43-104,90-95/Disulfide bonds: #status experimental

Query Match 55.3%; Score 94; DB 1; Length 105;

Best Local Similarity 67.9%; Pred. No. 3.4e-06;

Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29

Db 26 VNOHLCGSHLVYALVLCGERGFFYTPK 53

RESULT 24

IPDG

Insulin precursor - dog

C:Species: Canis lupus familiaris (dog)

C:Date: 24-Apr-1984 #sequence_revision 15-Nov-1984 #text_change 16-Jul-1999

C:Accession: A92413; A01587; S16493

R:Kwok, S.C.M.; Chan, S.J.; Steiner, D.F.

J. Biol. Chem. 258, 2357-2363, 1983

A:Title: Cloning and nucleotide sequence analysis of the dog insulin gene. Coded

A:Reference number: A92413; MUID:83109071; PMID:6296142

A:Accession: A92413

A:Molecule type: DNA

A:Residues: 1-110 <SMT>

A:Cross-references: GB:V00179; GB:J00042; NID:g994; PIDN:CAA23475.1; PID:g995

R:Smith, L.F.

Am. J. Med. 40, 662-666, 1966

A:Title: Species variation in the amino acid sequence of insulin.

A:Reference number: A90029; MUID:66160119; PMID:5949593

A:Accession: A01587

A:Molecule type: protein

A:Residues: 25-54;90-110 <SMIT>
R:Peterson, J.D.; Nehrllich, S.; Over, P.E.; Steiner, D.F.
J. Biol. Chem. 247, 4866-4871, 1972
A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsulin
A:Reference number: A92111; MUID:72258016; PMID:4626369
A:Accession: S16493

A:Molecule type: protein
A:Residues: 65-85, 'I', '87 <PET>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54/Domain: insulin chain B #status experimental <BCH>
F:25-54,90-110/Product: insulin #status experimental <MAT>
F:57-87/Domain: connecting peptide #status predicted <CPEP>
F:90-110/Domain: insulin chain A #status experimental <ACH>
F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 55.3%; Score 94; DB 1; Length 110;
Best Local Similarity 67.9%; Pred. No. 3.6e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 VNFHLYGSHIREWLYLVAGERGEFDPK 29
|| || ||| : |||| |||| : ||
Db 26 VNHLCGSHLVYALVCGERGEFFTPK 53

RESULT 25

INMKSQ

Insulin - common squirrel monkey

C:Species: Saimiri sciureus (common squirrel monkey)

C:Date: 18-Oct-1991 #sequence_revision 02-May-1994 #text_change 16-Jul-1999

C:Accession: B39258; A39258

F:Yu, J.H.; Eng, J.; Yalow, R.S.

Proc. Natl. Acad. Sci. U.S.A. 87, 9766-9768, 1990

A:Title: Isolation and amino acid sequences of squirrel monkey (Saimiri sciurea) insulin

A:Reference number: A39258; MUID:91088593; PMID:2263627

A:Accession: B39258

A:Molecule type: protein

A:Residues: 1-30 <YUB>

A:Accession: A39258

A:Molecule type: protein

A:Residues: 31-51 <YUA>

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:1-30,31-51/Product: insulin #status experimental <MAT>

F:31-51/Domain: insulin chain A #status experimental <ACH>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 54.7%; Score 93; DB 1; Length 51;
Best Local Similarity 65.5%; Pred. No. 2.2e-06;
Matches 19; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 2 VNFHLYGSHIREWLYLVAGERGEFDPK 30
|| || ||| : |||| |||| : |||
Db 2 VNHLCGPHLVYALVCGERGEFFYAPK 30

Search completed: December 23, 2002, 07:16:12
Job time : 15.2941 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:17 ; Search time 8.82353 Seconds
(without alignments)
141.020 Million cell updates/sec

Title: US-09-574-443-7_COPY_22_51

Perfect score: 170

Sequence: 1 DVNFHLYGSHIREWLYLVAGERGFNDPKT 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	99	58.2	51	INS_ELEMA	P01316 elephas max
2	99	58.2	110	INS_CERAE	P30407 cercopithec
3	99	58.2	110	INS_HUMAN	P01308 homo sapien
4	99	58.2	110	INS_MACFA	P30406 macaca fasc
5	99	58.2	110	INS_PANTR	P30410 pan troglod
6	96	56.5	51	INS_ANSAN	P07454 anser anser
7	96	56.5	81	INS_ANAPL	P01333 anas platyr
8	95	55.9	103	INS_SELRF	P51463 selasphorus
9	95	55.9	110	INS_CRILLO	P01313 cricetus
10	95	55.9	110	INS_RABIT	P01311 oryctolagus
11	94	55.3	51	INS_BALRO	P01314 balaeopter
12	94	55.3	51	INS_BALPH	P01312 balaeopter
13	94	55.3	51	INS_CAPHI	P01319 capra hircu
14	94	55.3	51	INS_DIDMA	P18109 didelphis m
15	94	55.3	51	INS_FELCA	P06306 felis silve
16	94	55.3	86	INS_HORSE	P01310 equus cabal
17	94	55.3	105	INS_BOVIN	P01317 bos taurus
18	94	55.3	105	INS_SHEEP	P01318 ovis aries
19	94	55.3	108	INS_PIG	P01315 sus scrofa
20	94	55.3	110	INS_PSAOB	P01321 canis famil
21	94	55.3	110	INS_PSAOB	Q62587 psammomys o
22	93	54.7	108	INS_AOTVR	P10604 aotus trivi
23	92	54.1	51	INS_ACOCA	P01324 aconyx cabi
24	91	53.5	51	INS_TRASC	P31887 trachemys s
25	91	53.5	107	INS_CHICK	P01332 gallus gall
26	90	52.9	50	INS_KATPE	P01340 katuswonus
27	90	52.9	51	INS_CAMDR	P01320 camelus dro
28	89	52.4	106	INS1_XENLA	P12706 xenopus lae
29	85	50.0	52	INS_ACIQU	P81423 acipenser g
30	85	50.0	106	INS2_XENLA	P12707 xenopus lae
31	84	49.4	52	INS_LEPSP	P09476 lepisosteus
32	84	49.4	108	INS1_MOUSE	P01325 mus musculu
33	84	49.4	110	INS1_RAT	P01322 rattus norv

34 83 48.8 50 1 INS_MYOSC
35 83 48.8 50 1 INS_ONCGO
36 83 48.8 51 1 INS1_BATSP
37 83 48.8 51 1 INS2_THUTH
38 83 48.8 51 1 INS_CHIBR
39 83 48.8 51 1 INS_GADCA
40 83 48.8 51 1 INS_ORNAN
41 83 48.8 110 1 INS2_MOUSE
42 83 48.8 110 1 INS2_RAT
43 83 48.8 116 1 INS_LOPPI
44 82 48.2 51 1 INS_ANGRO
45 81 47.6 51 1 INS_HYSOR

ALIGNMENTS

RESULT 1
INS_ELEMA STANDARD; PRT; 51 AA.
AC P01316;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Elephas maximus (Indian elephant).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.
OX NCBI_TaxID-9783;
RN [1]
RP SEQUENCE.
RX MEDLINE-66160119; PubMed-5949593;
RA Smith L.F.;
RT *Species variation in the amino acid sequence of insulin.*;
RL Am. J. Med. 40:662-666(1966).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: THE SPECIES OF ELEPHANT IS NOT GIVEN, BUT IT IS
CC MOST PROBABLY THE INDIAN ELEPHANT (ELEPHAS MAXIMUS).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
PIR: A01584; INEL.
HSSP: P01308; LA10.
InterPro: IPR004825; Ins/IGF/relax.
PRINTS: PR00276; INSULINA.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; IIGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5752 MW; 9007B50CDB457D6D CRC64;

Query Match 58.2%; Score 99; DB 1; Length 51;

Best Local Similarity 69.0%; Pred. No. 1:2e-07;

Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Oy 2 VNFHLYGSHIREWLYLVAGERGFNDPKT 30

||| ||| : ||| ||| : |||

Db 2 VNHLCGSHLVREALYLVCGERGFYTPKT 30

RESULT 2

INS_CERAE

```
ID INS_CERAE STANDARD; PRT; 110 AA.
AC P30407; P01309;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP MEDLINE-92219953; PubMed=1560757;
RA Seleno S., Bell G.I., Li W.;
RT "Sequences of primate insulin genes support the hypothesis of a
RT slower rate of molecular evolution in humans and apes than in
RT monkeys."
RL Mol. Biol. Evol. 9:193-203(1992).
RN [2]
RP SEQUENCE OF 57-87. PubMed=4626369;
RX MEDLINE-72258016; PubMed=4626369;
RA Peterson J.D., Neirlich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
RT dog proinsulin C-peptides by a semi-micro Edman degradation
RT procedure."
RL J. Biol. Chem. 247:4866-4871(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X61092; CAA43405.1; --
DR PIR; A05232; A05232.
DR PIR; S22056; S22056.
DR PIR; B42179; B42179.
DR HSP; P01308; IAI0.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;
Query Match 58.2%; Score 99; DB 1; Length 110;
Best Local Similarity 69.0%; Pred. No. 2.6e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
OY 2 VNFHLYGSHIREWLYLVAGERGFNFDPKT 30
Db 26 VNOHLCGSHLVYALVCGRGGFYTPKT 54
```

```
RESULT 3
INS_HUMAN STANDARD; PRT; 110 AA.
AC P01308;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Insulin precursor.
GN INS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-80120725; PubMed=6243748;
RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischer E.,
RA Goodman H.M.;
RT "Sequence of the human insulin gene."
RL Nature 284:26-32(1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-80236313; PubMed=6248962;
RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
RT "Genetic variation in the human insulin gene."
RL Science 209:612-615(1980).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-80054779; PubMed=503234;
RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
RA Rutter W.J.;
RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin."
RL Nature 282:525-527(1979).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-80147417; PubMed=6927840;
RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
RT "Nucleotide sequence of human preproinsulin complementary DNA."
RL Science 208:57-59(1980).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE-93364428; PubMed=8358440;
RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;
RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
RT kb segment of DNA spanning the insulin gene and associated VNTR."
RL Nat. Genet. 4:305-310(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE-Pancreas;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1-59 FROM N.A.
RC TISSUE-Blood;
RA Fajardy I.I., Weill J.J., Stuckens C.C., Danze P.M.P.;
RT "Description of a novel RFLP diallelic polymorphism (-127 BsgI C/G)
RT within the 5' region of insulin gene."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 25-54 AND 90-110.
RA Nicol D.S.H.W., Smith L.F.;
RT "Amino-acid sequence of human insulin."
RL Nature 187:483-485(1960).
RN [9]
RP SEQUENCE OF 57-87.
RX MEDLINE-71116410; PubMed=5101771;
RA Oyer P.E., Cho S., Peterson J.D., Steiner D.F.;
RT "Studies on human proinsulin. Isolation and amino acid sequence of
RT the human pancreatic C-peptide."
RL J. Biol. Chem. 246:1375-1386(1971).
RN [10]
```

RP SEQUENCE OF 57-87.
RX MEDLINE=71257722; PubMed=5560404;
RA KO A., Smyth D.G., Markussen J., Sundby F.;
RT "the amino acid sequence of the C-peptide of human proinsulin.";
RL Eur. J. Biochem. 20:190-199(1971).
RN [11]
RP SYNTHESIS.
RX MEDLINE=75077277; PubMed=4443293;
RA Sieber P., Kamber B., Hartmann A., Joehl A., Riniker B., Rittel W.;
RT "Total synthesis of human insulin under directed formation of the
RT disulfide bonds.";
RL Helv. Chim. Acta 57:2617-2621(1974).
RN [12]
RP SYNTHESIS OF 57-87.
RX MEDLINE=75040007; PubMed=4803504;
RA Naithani V.K.;
RT "Studies on polypeptides, IV. The synthesis of C-peptide of human
RT proinsulin.";
RL Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).
RN [13]
RP SYNTHESIS OF 65-69 AND 70-73.
RX MEDLINE=73161263; PubMed=4698555;
RA Gaiger R., Volk A.;
RT "Synthesis of peptides with the properties of human proinsulin C
RT peptides (hc peptide). 3. Synthesis of the sequences 14-17 and 9-13
RT of human proinsulin C peptides.";
RL Chem. Ber. 106:199-205(1973).
RN [14]
RP SYNTHESIS OF 84-87.
RX MEDLINE=73161261; PubMed=4698553;
RA Gaiger R., Jaeger G., Keonig W., Treuth G.;
RT "Synthesis of peptides with the properties of human proinsulin C
RT peptides (hc peptide). I. Scheme for the synthesis and preparation of
RT the sequence 28-31 of human proinsulin C peptide.";
RL Chem. Ber. 106:188-192(1973).
RN [15]
RP VARIANT LOS ANGELES.
RX MEDLINE=84016053; PubMed=6312455;
RA Haneada M., Chan S.J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;
RT "Studies on mutant human insulin genes: identification and sequence
RT analysis of a gene encoding [SerB24]insulin.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:6366-6370(1983).
RN [16]
RP VARIANTS LOS ANGELES AND CHICAGO.
RX MEDLINE=84170233; PubMed=6424111;
RA Shoelson S., Fickova M., Haneada M., Nahum G., Kaiser E.T.,
RA Rubenstein A.H., Tager H.;
RT "Identification of a mutant human insulin predicted to contain a
RT serine-for-phenylalanine substitution.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).
RN [17]
RP VARIANT PROVIDENCE.
RX MEDLINE=87175640; PubMed=3470784;
RA Chan S.J., Seino S., Gruppaso P.A., Schwartz R., Steiner D.F.;
RT "A mutation in the B chain coding region is associated with impaired
RT proinsulin conversion in a family with hyperproinsulinemia.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2194-2197(1987).
RN [18]
RP VARIANT WAKAYAMA.
RX MEDLINE=87056122; PubMed=3537011;
RA Sakura H., Iwamoto Y., Sakamoto Y., Kuzuya T., Hirata H.;
RT "Structurally abnormal insulin in a diabetic patient. Characterization
RT of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";
RL J. Clin. Invest. 78:1666-1672(1986).
RN [19]
RP VARIANT HIS-89.
RX MEDLINE=90317021; PubMed=2196279;
RA Barbeti F., Raben N., Kadowaki T., Cama A., Accili D., Gabbay K.H.,
RA Merenich J.A., Taylor S.I., Roth J.;
RT "Two unrelated patients with familial hyperproinsulinemia due to a
RT mutation substituting histidine for arginine at position 65 in the
RT proinsulin molecule: identification of the mutation by direct
RT sequencing of genomic deoxyribonucleic acid amplified by polymerase

RT chain reaction.";
RL J. Clin. Endocrinol. Metab. 71:164-169(1990).
RN [20]
RP VARIANT HIS-89.
RX MEDLINE=85261996; PubMed=4019786;
RA Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;
RT "Posttranslational cleavage of proinsulin is blocked by a point
RT mutation in familial hyperproinsulinemia.";
RL J. Clin. Invest. 76:378-380(1985).
RN [21]
RP VARIANT KYOTO.
RX MEDLINE=92291307; PubMed=1601997;
RA Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;
RT "A novel point mutation in the human insulin gene giving rise to
RT hyperproinsulinemia (proinsulin Kyoto).";
RL J. Clin. Invest. 89:1902-1907(1992).
RN [22]
RP STRUCTURE BY NMR.
RX MEDLINE=91104966; PubMed=2271664;
RA Hua Q.-X., Weiss M.A.;
RT "Toward the solution structure of human insulin: sequential 2D 1H NMR
RT assignment of a des-pentapeptide analogue and comparison with crystal
RT structure.";
RL Biochemistry 29:10545-10555(1990).
RN [23]
RP STRUCTURE BY NMR.
RX MEDLINE=91242467; PubMed=2036420;
RA Hua Q.-X., Weiss M.A.;
RT "Comparative 2D NMR studies of human insulin and des-pentapeptide
RT insulin: sequential resonance assignment and implications for protein
RT dynamics and receptor recognition.";
RL Biochemistry 30:5505-5515(1991).
RN [24]
RP STRUCTURE BY NMR.
RX MEDLINE=91265527; PubMed=1646635;
RA Hua Q.-X., Weiss M.A.;
RT "Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-
RT specific resonance assignments and effects of solvent composition.";
RL Biochim. Biophys. Acta 1078:101-110(1991).
RN [25]
RP STRUCTURE BY NMR.
RX MEDLINE=93059366; PubMed=1433291;
RA Joergensen A.M.M., Kristensen S.M., Led J.J., Balschmidt P.;
RT "Three-dimensional solution structure of an insulin dimer. A study of
RT the B9(Asp) mutant of human insulin using nuclear magnetic resonance,
RT distance geometry and restrained molecular dynamics.";
RL J. Mol. Biol. 227:1146-1163(1992).
RN [26]
RP STRUCTURE BY NMR OF VARIANT LOS-ANGELES.
RX MEDLINE=93133832; PubMed=8421693;
RA Hua Q.-X., Shoelson S.E., Inouye K., Weiss M.A.;
RT "Paradoxical structure and function in a mutant human insulin
RT associated with diabetes mellitus.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:582-586(1993).
RN [27]
RP STRUCTURE BY NMR.
RX MEDLINE=97383146; PubMed=9235985;
RA Chang X., Joergensen A.M., Bardrum P., Led J.J.;
RT Query Match 58.2%; Score 99; DB 1; Length 110;
RT Best Local Similarity 69.0%; Pred. No. 2.6e-07;
RT Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
D6 26 VNOHLCGSHLVEALYLVCGERGFFYTPKT 54
RESULT 4
ID INS_MACFA STANDARD; PRT; 110 AA.
AC P30406; P01309;
DT 21-JUL-1986 (Rel. 01, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DE 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Insulin precursor.
 INS..
 GN Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=83080474; PubMed=6184262;
 RX Winkler W., Gronenberg J., Lelneweber M., Wengenmayer F.,
 RA "The nucleotide sequence of cDNA coding for preproinsulin from the
 RT primate Macaca fascicularis.";
 RL Gene 19:179-183(1982).
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J00336; AAA36849.1; -
 DR PIR; JQ0178; JQ0178.
 DR HSP; P01308; IAI0.
 DR InterPro: IPR004825; Ins/IGF/relax.
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS; PR00276; INSULIN.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 DR Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INSULIN A CHAIN.
 FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100 INTERCHAIN.
 SQ SEQUENCE 110 AA; 11991 MW; 83C6E833A80A20F9 CRC64;
 Query Match 58.2%; Score 99; DB 1; Length 110;
 Best Local Similarity 69.0%; Pred. No. 2.6e-07;
 Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 VNFHLYGSHIREWLYVAGERGFNDPKT 30
 DB 26 VNOHLCGSHLVEALYLVCGERGFFTPKT 54
 RESULT 5
 INS_PANTR STANDARD; PRT; 110 AA.
 AC P30410;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Insulin precursor.
 GN INS.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92219953; PubMed=1560757;
 RX Seino S., Bell G.I., Li W.;
 RA "Sequences of primate insulin genes support the hypothesis of a
 RT slower rate of molecular evolution in humans and apes than in
 RT monkeys.";
 RL Mol. Biol. Evol. 9:193-203(1992).
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X61089; CAA43403.1; -
 DR PIR; S22058; S22058.
 DR PIR; A42179; A42179.
 DR HSP; P01308; IAI0.
 DR InterPro: IPR004825; Ins/IGF/relax.
 DR Pfam: PF00049; Insulin; 1.
 DR PRINTS; PR00276; INSULIN.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 DR Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INSULIN A CHAIN.
 FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100 INTERCHAIN.
 SQ SEQUENCE 110 AA; 12025 MW; 41EB8DF79837CEF5 CRC64;
 Query Match 58.2%; Score 99; DB 1; Length 110;
 Best Local Similarity 69.0%; Pred. No. 2.6e-07;
 Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 2 VNFHLYGSHIREWLYVAGERGFNDPKT 30
 DB 26 VNOHLCGSHLVEALYLVCGERGFFTPKT 54
 RESULT 6
 INS_ANGAN STANDARD; PRT; 51 AA.
 AC P07454; Q10995;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Insulin.
 GN INS.
 OS Anser anser anser (Western graylag goose), and
 OS Cairina moschata (Muscovy duck).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anser.
 OX NCBI_TaxID=8844, 8855;
 RN [1]
 RN SEQUENCE.
 RP SPECIES=A.anser;

CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
-!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
-!- SUBCELLULAR LOCATION: Secreted.
-!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
-!- CAUTION: Y/S AT POSITIONS 31-32 AND 59-60 REPRESENT PAIRED BASIC RESIDUES ASSUMED BY HOMOLOGY TO BE PRESENT IN THE PRECURSOR MOLECULE.

PIR; A01600; IPDK.
DR HSSP; P01308; LHIS.
DR Pfam; PR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
CC CHAIN 1 30 INSULIN B CHAIN.
FT FT PROPEP 33 58 C PEPTIDE.
FT CHAIN 61 81 INSULIN A CHAIN.
FT DISULFID 7 67 INTERCHAIN.
FT DISULFID 19 80 INTERCHAIN.
FT DISULFID 66 71
CC SEQUENCE 81 AA; 9100 MW; 6EA8A271F099DA91 CRC64;

Query Match 56.5%; Score 96; DB 1; Length 81;
Best Local Similarity 67.9%; Pred. No. 5.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 NFHLGSHIREWLYLVAGERGCFNFDPKT 30
| | |||| : |||| |||| : ||||
DB 3 NQLCGSHLVEALYLVCGERGFFYSPTK 30

RESULT 8
INS_SELF
ID INS_SELF STANDARD; PRT; 103 AA.
AC F51463;
DT DT 01-OCT-1996 (Rel. 34, Created)
DT DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Insulin precursor (Fragment).
GN INS.
OS Salsaporphus rufus (Hummingbird).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauia; Aves; Neognathae; Trochiliformes; Trochilidae;
OC Salsaporphus.
OX NCBI_TaxID=29060;
RN [1]
RZ SEQUENCE FROM N.A.
RX MEDLINE=94010162; PubMed=8405887;
FA Pan L., Gardner P., Chan S.J., Steiner D.F.;
RA "Cloning and analysis of the gene encoding hummingbird proinsulin.";
RL Gen. Comp. Endocrinol. 91:25-30(1993).

-!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
-!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
-!- SUBCELLULAR LOCATION: Secreted.
-!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to licenset@isb-sib.ch).

EMBL; S66612; AAC64211.1; -
DR EMBL; S66611; AAC64211.1; JOINED.

DR HSSP; P01308; IHIS.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; IIGF; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 1
FT NON_TER 1 1
FT CHAIN <1 20 BY SIMILARITY.
FT CHAIN 21 50 INSULIN B CHAIN.
FT PROPEP 53 80 C PEPTIDE.
FT CHAIN 83 103 INSULIN A CHAIN.
FT CHAIN 83 103 INSULIN A CHAIN.
FT DISULFID 27 89 INTERCHAIN (BY SIMILARITY).
FT DISULFID 39 102 INTERCHAIN (BY SIMILARITY).
FT DISULFID 88 93 BY SIMILARITY.
SQ SEQUENCE 103 AA; 11378 MW; 6598520B28F5BB9C CRC64;

Query Match 55.98; Score 95; DB 1; Length 103;
Best Local Similarity 67.98; Pred. No. 9.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNFDPK 29
II III: I IIII IIIII: II
DB 22 VNOHLCGSHLVYALYVCGERGFYSPK 49

RESULT 9

INS_CRILLO STANDARD; PRT; 110 AA.
AC P01313;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84133036; PubMed=6365663;
RA Beil G.I., Sanchez-Pescador R.;
RT "Sequence of a cDNA encoding Syrian hamster preproinsulin.";
RL Diabetes 33:297-300(1984).
RN [2]
RP SEQUENCE OF 25-54 AND 90-110.
RA Neelon F.A., Delcher H.K., Steinman H., Lebovitz H.E.;
RT "Structure of hamster insulin: comparison with a tumor insulin.";
RL Fed. Proc. 32:300-300(1973).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M26328; AAA37089.1;
CC PIR; A91456; INHY.
CC HSSP; P01308; IYIM.
CC InterPro; IPR004825; Ins/IGF/relax.
CC Pfam; PF00049; Insulin; 1.
CC PRINTS; PR00276; INSULINA.

DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
SQ SEQUENCE 110 AA; 12268 MW; 219E92B85A535CEC CRC64;

Query Match 55.98; Score 95; DB 1; Length 110;
Best Local Similarity 65.5%; Pred. No. 9.7e-07;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNFDPK 30
II III: I IIII IIIII: II
DB 26 VNOHLCGSHLVYALYVCGERGFYTPKS 54

RESULT 10

INS_RABIT STANDARD; PRT; 110 AA.
AC P01311;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin precursor.
GN INS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-New Zealand white; TISSUE-Pancreas;
RX MEDLINE=94179230; PubMed=8132571;
RA Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R.,
RA Menon R.K., Zahm D.S.;
RT "Insulin gene expression and insulin synthesis in mammalian neuronal
RT cells.";
RL J. Biol. Chem. 269:8445-8454(1994).
RN [2]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.";
RL Am. J. Med. 40:662-666(1966).
RN [3]
RP SEQUENCE OF 56-110 FROM N.A.
RA Giddings S.J., Carnaghi L.R., Devaskar S.U.;
RC Submitted (APR-1991) to the EMBL/GenBank/DDJ databases.
RL -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U03610; AAA19033.1;
CC EMBL; M61153; AAA17540.1;

DR PIR: A01581; INRB.
DR HSP; P01308; IRYM.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INSULIN A CHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100
FT CONFLICT 83
SQ SEQUENCE 110 AA; 11838 MW; 82D2975B85D77FA8 CRC64;

Query Match 55.9%; Score 95; DB 1; Length 110;
Best Local Similarity 65.5%; Pred. No. 9.7e-07;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNFDPKT 30
DB 26 VNOHLCGSHLVEALYLVCGERGEFFTYPKS 54

RESULT 11
INS_BALBO STANDARD; PRT; 51 AA.
AC P01314;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin.
GN INS.
OS Balaenoptera borealis (Sei whale).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9768;
RN [1]
RP SEQUENCE.
RA Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by
RT whale pepsin.";
RL Nature 181:1468-1469(1958).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01582; INWHF.
DR HSP; P01317; IAPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;

Query Match 55.3%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 6.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Best Local Similarity 67.9%; Pred. No. 6.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNFDPK 29
DB 2 VNOHLCGSHLVEALYLVCGERGEFFTYPK 29

RESULT 12
INS_BALPH STANDARD; PRT; 51 AA.
AC P01312;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin.
GN INS.
OS Balaenoptera physalus (Finback whale) (Common rorqual), and
OS Physeter catodon (Sperm whale) (Physeter macrocephalus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9770, 9755;
RN [1]
RP PARTIAL SEQUENCE.
RA Hama H., Titani K., Sakaki S., Narita K.;
RT "The amino acid sequence in fin-whale insulin.";
RL J. Biochem. 56:285-293(1964).
RN [2]
RP SEQUENCE.
RC SPECIES-P.catodon;
RA Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by
RT whale pepsin.";
RL Nature 181:1468-1469(1958).
RN [3]
RP SEQUENCE.
RC SPECIES-P.catodon;
RA Harris J.I., Sanger F., Naughton M.A.;
RT "Species differences in insulin.";
RL Arch. Biochem. Biophys. 65:427-438(1956).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A91918; INWHF.
DR PIR: A91142; INWHP.
DR HSP; P01317; IAPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5766 MW; 9007B514691A7CDD CRC64;

Query Match 55.3%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 6.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNFDPK 29
DB 2 VNOHLCGSHLVEALYLVCGERGEFFTYPK 29

```
Db 2 VNOHLCGSHLVREALYLVCGERGFFFTPK 29

RESULT 13
INS_CAPHI STANDARD; PRT; 51 AA.
AC P01319;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.";
RL Am. J. Med. 40:662-666(1966).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC PIR; A01586; INGT.
DR HSSP; P01317; IAPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5692 MW; 9007B50CDB4E7DDD CRC64;

Query Match 55.3%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 6.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNEDPK 29
DB 2 VNOHLCGSHLVREALYLVCGERGFFFTPK 29

RESULT 14
INS_DIDMA STANDARD; PRT; 51 AA.
AC P18109;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Didelphis marsupialis virginiana (North American opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.
OX NCBI_TaxID=9267;
RN [1]
RP SEQUENCE.
RX TISSUE=Pancreas;
RX MEDLINE=90160042; PubMed=2695899;

Yu J.-H., Eng J., Rattan S., Yalow R.S.;
"Opossum insulin, glucagon and pancreatic polypeptide: amino acid
sequences.";
Peptides 10:1195-1197(1989).
-1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
-1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
DISULFIDE BONDS.
-1- SUBCELLULAR LOCATION: Secreted.
-1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
PIR; JQ0362; JQ0362.
DR PIR; JQ0363; JQ0363.
DR HSSP; P01317; 2INS.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5732 MW; 9007B8BAE4BDEEDD CRC64;

Query Match 55.3%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 6.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNEDPK 29
DB 2 VNOHLCGSHLVREALYLVCGERGFFFTPK 29

RESULT 15
INS_FELCA STANDARD; PRT; 51 AA.
AC P06306;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin.
GN INS.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE.
RX MEDLINE=86214076; PubMed=3518635;
RA Halliday G., Gavellin G., Mutt V., Joernvall H.;
"Characterization of cat insulin.";
Arch. Biochem. Biophys. 247:20-27(1986).
-1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
-1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
DISULFIDE BONDS.
-1- SUBCELLULAR LOCATION: Secreted.
-1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
PIR; A01588; INCT.
DR HSSP; P01317; IAPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
```

FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5745 MW; 9007B5096A0A7DDD CRC64;

Query Match 55.3%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 6.1e-07;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNEDPK 29
II III: I IIII IIIII: II
Db 2 VNHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 16
INS_HORSE STANDARD; PRT; 86 AA.
AC P01310;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Insulin precursor.
GN INS.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE OF 1-30 AND 66-86.
RA Harris J. I., Sanger F., Naughton M. A.;
RT "Species differences in insulin.";
RL Arch. Biochem. Biophys. 65:427-438(1956).
RN [2]
RP SEQUENCE OF 33-63.
RX MEDLINE-73061498; PubMed-4640931;
RA Tager H. S., Steiner D. F.;
RT "Primary structures of the proinsulin connecting peptides of the rat and the horse.";
RL J. Biol. Chem. 247:7936-7940(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -1- CAUTION: X'S AT POSITIONS 31-32 AND 64-65 REPRESENT PAIRED BASIC RESIDUES ASSUMED BY HOMOLGY TO BE PRESENT IN THE PRECURSOR MOLECULE.
DR PIR; A01580; IPHO.
DR HSSP; P01317; IAPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF000049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SH00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT PROPEP 33 63 C PEPTIDE.
FT CHAIN 66 86 INSULIN A CHAIN.
FT DISULFID 7 72 INTERCHAIN.
FT DISULFID 19 85 INTERCHAIN.
FT DISULFID 71 76
SQ SEQUENCE 86 AA; 9142 MW; A3E1E822711BDB46 CRC64;

Query Match 55.3%; Score 94; DB 1; Length 86;
Best Local Similarity 67.9%; Pred. No. 1e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNEDPK 29
II III: I IIII IIIII: II
Db 2 VNHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 17
INS_BOVIN

ID INS_BOVIN STANDARD; PRT; 105 AA.
AC P01317;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin precursor.
GN INS.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-88288209; PubMed-2456452;
RA D'Agostino J., Younes M. A., White J. W., Besch P. K., Field J. B., Frazier M. L.;
RT "Cloning and nucleotide sequence analysis of complementary deoxyribonucleic acid for bovine preproinsulin.";
RL Mol. Endocrinol. 1:327-331(1987).
RN [2]
RP SEQUENCE OF 25-105.
RX MEDLINE-71166442; PubMed-4928892;
RA Nolan C., Margoliash E., Peterson J. D., Steiner D. F.;
RT "The structure of bovine proinsulin.";
RL J. Biol. Chem. 246:2780-2795(1971).
RN [3]
RP SEQUENCE OF 25-54.
RA Sanger F., Tuppy H.;
RT "The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates.";
RL Biochem. J. 49:481-490(1951).
RN [4]
RP SEQUENCE OF 57-82.
RX MEDLINE-71116409; PubMed-5545080;
RA Steiner D. F., Cho S., Oyer P. E., Terris S., Peterson J. D., Rubenstein A. H.;
RT "Isolation and characterization of proinsulin C-peptide from bovine pancreas.";
RL J. Biol. Chem. 246:1365-1374(1971).
RN [5]
RP SEQUENCE OF 57-82.
RX MEDLINE-71257721; PubMed-5105368;
RA Salokangas A., Smyth D. G., Markussen J., Sundby F.;
RT "Bovine proinsulin: amino acid sequence of the C-peptide isolated from pancreas.";
RL Eur. J. Biochem. 20:183-189(1971).
RN [6]
RP SEQUENCE OF 85-105.
RA Sanger F., Thompson E. O. P.;
RT "The amino-acid sequence in the glycyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates.";
RL Biochem. J. 53:366-374(1953).
RN [7]
RP AMIDES, SEQUENCE OF 25-54 AND 85-105, AND DISULFIDE BONDS.
RA Kyle A. P., Sanger F., Smith L. F., Kitai R.;
RT "The disulphide bonds of insulin.";
RL Biochem. J. 60:541-556(1955).
RN [8]
RP X-RAY CRYSTALLOGRAPHY.
RA Smith G. D., Duax W. L., Dodson E. J., Dodson G. G., de Graaf R. A. G., Reynolds C. D.;
RT "The structure of des-Phe b1 bovine insulin.";
RL Acta Crystallogr. B 38:3028-3032(1982).
RN [9]

RP X-RAY CRYSTALLOGRAPHY (1.3 ANGSTROMS):
RX MEDLINE-97285914; PubMed-9141131;
RA Brange J., Dodson G.G., Edwards D.J., Holden P.H., Whittingham J.L.;
RT "A model of insulin fibrils derived from the X-ray crystal structure
of a monomeric insulin (despentapeptide insulin).";
RL Proteins 27:507-516(1997).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -!- DATABASE: NAME-Protein Spotlight;
CC NOTE-Issue 9 of April 2001;
CC WWW="http://www.expasy.org/spotlight/articles/sptlt009.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M54979; AAA30722.1; -
DR PIR; A01585; IPBO.
DR PIR; A40909; A40909.
DR PDB; 2INS; 31-MAY-84.
DR PDB; 1APH; 31-OCT-93.
DR PDB; 1BPH; 31-OCT-93.
DR PDB; 1CPH; 31-OCT-93.
DR PDB; 1DPH; 31-OCT-93.
DR PDB; 1PID; 07-DEC-96.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
DR Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 82 C PEPTIDE.
FT CHAIN 85 105 INSULIN A CHAIN.
FT DISULFID 31 91 INTERCHAIN.
FT DISULFID 43 104
FT TURN 32 32
FT TURN 33 46
FT STRAND 48 48
FT HELIX 86 90
FT TURN 91 94
FT HELIX 97 101
FT TURN 102 103
FT STRAND 104 104
SQ SEQUENCE 105 AA; 11393 MW; 75307CF78E61C06A CRC64;

Query Match 55.3%; Score 94; DB 1; Length 105;
Best Local Similarity 67.9%; Pred. No. 1.3e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
|| || |||: | |||| ||||| : ||
DB 26 VNOHLCGSHLVEALYLVCGERGFFYTPK 53

RESULT 18
INS_SHEEP STANDARD; PRT; 105 AA.
AC P01318;
DT 21-JUL-1986 (Rel. 01, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin precursor.
GN INS.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94280618; PubMed-8011164;
RA Ohlssen S.M., Lugenbeel K.A., Wong E.A.;
RT "Characterization of the linked ovine insulin and insulin-like growth
factor-II genes.";
RL DNA Cell Biol. 13:377-388(1994).
RN [2]
RP SEQUENCE OF 25-54 AND 85-105.
RA Brown H., Sanger F., Kitai R.;
RT "The structure of pig and sheep insulins.";
RL Biochem. J. 60:556-565(1955).
RN [3]
RP SEQUENCE OF 57-82.
RX MEDLINE-72258016; PubMed-4626369;
RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
dog proinsulin C-peptides by a semi-micro Edman degradation
procedure.";
RL J. Biol. Chem. 247:4866-4871(1972).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00659; AAB60625.1; -
DR PIR; S16430; INSH.
DR HSSP; P01317; IAPH.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
DR Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 82 C PEPTIDE.
FT CHAIN 85 105 INSULIN A CHAIN.
FT DISULFID 31 91 INTERCHAIN.
FT DISULFID 43 104
FT DISULFID 90 95
SQ SEQUENCE 105 AA; 11235 MW; 8B27C7FB9922BC7A CRC64;

Query Match 55.3%; Score 94; DB 1; Length 105;
Best Local Similarity 67.9%; Pred. No. 1.3e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
|| || |||: | |||| ||||| : ||
DB 26 VNOHLCGSHLVEALYLVCGERGFFYTPK 53

```
RESULT 19
INS_PIG
ID INS_PIG STANDARD; PRT; 108 AA.
AC P01315; Q9TSJ5;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin precursor.
GN INS.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Han X.G., Tuch B.E.;
RA "Complete porcine preproinsulin cDNA sequence.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=68286485; PubMed=5657063;
RA Chance R.E., Ellis R.M., Bromer W.W.;
RT "Porcine proinsulin: characterization and amino acid sequence.";
RL Science 161:165-167(1968).
RN [3]
RP REVISION TO 59.
RA Chance R.E.;
RL Submitted (JUL-1970) to the PIR data bank.
RN [4]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;
RT "Insulin. The structure in the crystal and its reflection in
chemistry and biology.";
RL Adv. Protein Chem. 26:279-402(1972).
RN [5]
RX X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RA Isaacs N.W., Agarwal R.C.;
RT "Experience with fast Fourier least squares in the refinement of the
crystal structure of rhombohedral 2-zinc insulin at 1.5-A
resolution.";
RL Acta Crystallogr. A 34:782-791(1978).
RN [6]
RX X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX MEDLINE=89099318; PubMed=2905485;
RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J.,
RA Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W.,
RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;
RT "The structure of 2zn pig insulin crystals at 1.5-A resolution.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456(1988).
RN [7]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=92126280; PubMed=1772633;
RA Balschmidt P., Hansen F.B., Dodson E., Dodson G., Korber F.;
RT "Structure of porcine insulin cocrystallized with clupeine 2.";
RL Acta Crystallogr. B 47:975-986(1991).
RN [8]
RX X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=91222450; PubMed=2025410;
RA Badger J., Harris M.R., Reynolds C.D., Evans A.C., Dodson E.J.,
RA Dodson G.G., North A.C.T.;
RT "Structure of the pig insulin dimer in the cubic crystal.";
RL Acta Crystallogr. B 47:127-136(1991).
RN [9]
RX X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RA Diao J.-S., Wan Z.-L., Chang W.-R., Liang D.-C.;
RT "Structure of monomeric porcine DesB1-B2 desptapeptide (B36-B30)
insulin at 1.65-A resolution.";
RL Acta Crystallogr. D 53:507-512(1997).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
```

CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -1- DATABASE: NAME-Protein Spotlight;
CC NOTE-Issue 9 of April 2001;
CC WWW="http://www.expasy.org/spotlight/articles/sptit009.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF064555; AAC77920.1; ALT_INIT.
DR PIR; A01583; IPPG.
DR PDB; 3INS; 09-JAN-89.
DR PDB; 4INS; 31-JUL-94.
DR PDB; 6INS; 31-JAN-94.
DR PDB; 7INS; 31-JAN-94.
DR PDB; 9INS; 15-OCT-91.
DR PDB; 11ZA; 15-OCT-91.
DR PDB; 11ZB; 15-OCT-91.
DR PDB; 2TCI; 29-JAN-96.
DR PDB; 1MPJ; 29-JAN-96.
DR PDB; 3MTH; 29-JAN-96.
DR PDB; 1DEI; 16-JUN-97.
DR PDB; 1SDB; 01-APR-98.
DR PDB; 1WAV; 28-FEB-97.
DR PDB; 1ZEI; 16-FEB-99.
DR PDB; 1ZNI; 28-JAN-98.
DR PDB; 1ZNI; 28-JAN-98.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 85 C PEPTIDE.
FT CHAIN 88 108 INSULIN A CHAIN.
FT DISULFID 31 94 INTERCHAIN.
FT DISULFID 43 107 INTERCHAIN.
FT DISULFID 93 98
FT HELIX 26 46
FT STRAND 48 48
FT HELIX 89 94
FT HELIX 100 106
FT STRAND 107 107
SQ SEQUENCE 108 AA; 11671 MW; CB4491B429858EBE CRC64;

Query Match 55.3%; Score 94; DB 1; Length 108;
Best Local Similarity 67.9%; Pred. No. 1.3e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNDPK 29
||| |||: |||| ||||: ||
Db 26 VNHLCGSHLVEALYLVCGERGFFYTPK 53

RESULT 20
INS CANFA
ID INS_CANFA STANDARD; PRT; 110 AA.
AC P01321;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Psammomys.
OX NCBI_TaxID=48139;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=97309250; PubMed=9166665;
RA Kaiser N., Baillys E.M., Schneider B.S., Cerasi E., Steiner D.F.,
RA Hutton J.C., Gross D.J.;
RA *Characterization of the unusual insulin of *Psammomys obesus*, a
RP rodent with nutrition-induced NIDDM-like syndrome.";
RL Diabetes 46:953-957(1997).
RT -|- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLUCOGEN SYNTHESIS IN LIVER.
CC -|- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X98241; CAA66897.1; -;
DR HSP; P01308; IAI0.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINE.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24 BY SIMILARITY.
FT CHAIN 25 54 INSULIN B CHAIN..
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).
FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).
FT DISULFID 95 100 BY SIMILARITY.
SQ SEQUENCE 110 AA; 12324 MW; A006738E20579CB0 CRC64;

Query Match 55.3%; Score 94; DB 1; Length 110;
Best Local Similarity 67.9%; Pred. No. 1.3e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNFDPK 29
DB 26 VNOHLAGSHLVEALYLVCGERGFFYTPK 53
|| || ||| : |||| ||||| : ||
|| || ||| : |||| ||||| : ||

RESULT 22
INS_AOTTR INS_AOTTR STANDARD; PRT; 108 AA.
ID ID AC
AC AC
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin precursor.
GN GN
OS Aotus trivirgatus (Night monkey) (Douroucoulli), and
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotus.
OX NCBI_TaxID=9505; 9521;
RN [1]
RP SEQUENCE FROM N.A.

SPECIES-A. trivirgatus;
MEDLINE=88041119; PubMed=3118367;
Seino S., Steiner D.F., Bell G.I.;
"Sequence of a New World primate insulin having low biological
potency and immunoreactivity".
Proc. Natl. Acad. Sci. U.S.A. 84:7423-7427(1987).
[2]
SEQUENCE OF 25-54 AND 88-108.
SPECIES-S.sciureus;
MEDLINE=g1088593; PubMed=2263627;
Yu J.-H., Eng J., Yalow R.S.;
"Isolation and amino acid sequences of squirrel monkey (Saimiri
sciurea) insulin and glucagon".
Proc. Natl. Acad. Sci. U.S.A. 87:9766-9768(1990).
-I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
-I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
DISULFIDE BONDS.
-I- SUBCELLULAR LOCATION: Secreted.
-I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

EMBL	J02989	AAA35374.1	..	
PIR	B39258	INMKSQ.		
PIR	A39883	A39883.		
HSSP	P01308	1HIS.		
InterPro	IPR004825	Ins/IGF/relax.		
Pfam	PF00049	Insulin; 1.		
PRINTS	PRO0276	INSULINA.		
PRINTS	PRO0277	INSULINB.		
SMART	SM00078	ILGF; 1.		
PROSITE	PS00262	INSULIN; 1.		
Insulin family	Hormone	Glucose metabolism; Signal.		
SIGNAL	1	24		
CHAIN	25	54.	INSULIN B CHAIN.	
PROPEP	57	85	C PEPTIDE.	
CHAIN	88	108	INSULIN A CHAIN.	
DISULFID	31	94	INTERCHAIN.	
DISULFID	43	107	INTERCHAIN.	
DISULFID	93	98		
SEQUENCE	108 AA	11842 MW		1869B8250099731F CRC64;

	Query Match	54.7%;	Score 93;	DB 1;	Length 108;
	Best Local Similarity	65.5%;	Pred. No. 1.8e-06;		
	Matches 19;	Conservative	2;	Mismatches 8;	Indels 0;
	Gaps	0;			
QY	2	VNFHLYGSHIREWLVLVAGERGFNFDPKT	30		
DB	26	VNQHLCGPHLVLEALYLVCGERGFYAPKT	54		

RESULT 23			
ID	INS_ACOCA	STANDARD;	PRT; 51 AA.
AC	P01324;		
DT	21-JUL-1986	(Rel. 01, Created)	
DT	21-JUL-1986	(Rel. 01, Last sequence update)	
DT	01-NOV-1997	(Rel. 35, Last annotation update)	
DE	Insulin.		
DE	Insulin.		
GN	INS.		
OS	Acomys cahirinus (Egyptian spiny mouse).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Acomys.		
OX	NCBI TaxID-10068;		

```

[1]
RN      COMPOSITION.
RP      MEDLINE-72189454; PubMed-5028210;
RX      Buenzli H.F., Humbel R.E.;
RT      "Isolation and partial structural analysis of insulin from mouse (Mus
RT      musculus) and spiny mouse (Acomys cahirinus)".
RT      Hoppe-Seyler's Z. Physiol. Chem. 353:444-450(1972).
CC      -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC      INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC      FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC      CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC      -1- SUBUNIT: HETERO-DIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC      DISULFIDE BONDS.
CC      -1- SUBCELLULAR LOCATION: Secreted.
CC      -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR      PIR; A01591; INNSPP.
DR      HSSP; P01308; ITYM.
DR      InterPro; IPR004825; Ins/IGF/relax.
DR      PRINTS; PR00276; INSULINA.
DR      PRINTS; PR00277; INSULINB.
DR      SMART; SM00078; IIGF; 1.
DR      PROSITE; PS00262; INSULIN; 1.
KW      Insulin family; Hormone; Glucose metabolism.
FT      CHAIN      1      30      INSULIN B CHAIN.
FT      NON_CONS   30      31
FT      CHAIN      31      51      INSULIN A CHAIN.
FT      DISULFID   7      37      INTERCHAIN (BY SIMILARITY).
FT      DISULFID  19      50      INTERCHAIN (BY SIMILARITY).
FT      DISULFID  36      41      BY SIMILARITY.
FT      SEQUENCE   51 AA; 5768 MW; 992BD8B629047D3D CRC64;
SQ
Query Match          54.1%; Score 92; DB 1; Length 51;
Best Local Similarity 62.1%; Pred. No. 1.2e-06;
Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps
QY      2 VNFHLYGSHIREWLYLVAGERGCFNFDPKT 30
       1: 11 111: 1 1111 11111: 11:
DB      2 VBQHLGGSHLVYALVCGERGFFYTPKS 30
RESULT 24
INS_TRASC
ID      INS_TRASC      STANDARD;      PRT;      51 AA:
RC      P31887;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Insulin.
DE      INS.
GN      Trachemys scripta (Red-eared slider turtle) (Pseudemys scripta), and
OS      Chrysemys dorsigni (Black-bellied slider turtle) (Trachemys dorsigni)
OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Testudines; Cryptodira; Testudinoidea; Emydidae; Trachemys.
OX      NCBI_Taxid=34903, 31137;
[1]
RN      SEQUENCE.
RP      SPECIES-T. scripta;
RX      MEDLINE-90341082; PubMed-1974347;
RA      Conlon J.M., Hicks J.W.;
RT      "Isolation and structural characterization of insulin, glucagon and
RT      somatostatin from the turtle, Pseudemys scripta.";
RL      Peptides 11:461-466(1990).
[2]
RN      SEQUENCE.
RP      SPECIES-C. dorsigni;
RX      MEDLINE-92225302; PubMed-1808015;
RA      Cascone O., Turyn D., Dellacha J.M., Machado V.L.A., Marques M.,
RA      Vita N., Cassan C., Ferrara P., Guillemot J.C.;
RT      "Isolation, purification and primary structure of insulin from the
RT      turtle Chrysemys dorsigni.";
RL      Gen. Comp. Endocrinol. 84:355-359(1991).
CC      -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC      INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:17 ; Search time 29.4118 Seconds
(without alignments)
210.168 Million cell updates/sec

Title: us-09-574-443-7_COPY_22_51

Perfect score: 170

Sequence: 1 DVNFHLYGSHIREWLYLVAGERGFNFDPKT 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhch.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	95	55.9	110	11 Q91X13	Q91x13 spermophilu
2	94	55.3	110	6 Q8WNW6	Q8wnw6 felis silve
3	92	54.1	111	13 Q98TR0	Q98tb0 chitola chi
4	87	51.2	87	13 Q98TA9	Q98ta9 gnathonemus
5	85	50.0	110	13 Q98TA8	Q98ta8 pantodon bu
6	85	50.0	110	13 Q90Z11	Q90zyl hiodon alos
7	83	48.8	91	13 Q98TB2	Q98tb2 ambloplites
8	74	43.5	106	13 Q91BQ7	Q91bq7 rana pipien
9	74	43.5	108	13 Q98TB1	Q98tb1 catostomus
10	74	43.5	108	13 Q9DDP5	Q9dde5 brachydanio
11	74	43.5	108	13 Q90ZM4	Q90zn4 catla catla
12	73	42.9	111	13 Q98TA7	Q98ta7 osteoglossu
13	59	34.7	151	8 Q9TCV1	Q9tcv1 metagonimus
14	58.5	34.4	328	8 Q8SMK3	Q8smk3 ipomopsis m
15	58.5	34.4	358	8 Q32253	Q32253 gilia austr
16	58.5	34.4	358	8 Q32268	Q32268 gilia carui

17	58.5	34.4	358	8	Q32356	Q32356 gilia splen
18	57.5	33.8	360	8	Q32809	Q32809 polemonium
19	57.5	33.8	360	8	Q9THX7	Q9thx7 polemonium
20	57	33.5	148	8	Q9TCV2	Q9tcv2 metagonimus
21	57	33.5	154	17	Q8TZA8	Q8tza8 methanopyru
22	57	33.5	299	16	Q9KBM0	Q9kbm0 bacillus ha
23	56.5	33.2	358	8	Q32181	Q32181 eriastrum d
24	56.5	33.2	358	8	Q32287	Q32287 gilia filif
25	56.5	33.2	358	8	Q32290	Q32290 gilia hutch
26	56.5	33.2	358	8	Q32442	Q32442 ipomopsis a
27	56.5	33.2	358	8	Q32464	Q32464 ipomopsis p
28	56.5	33.2	358	8	Q32300	Q32300 gilia lepto
29	56.5	33.2	358	8	Q32538	Q32538 langloisia
30	56.5	33.2	358	8	Q32826	Q32826 phlox grac
31	56.5	33.2	358	8	Q9THY5	Q9thy5 acanthogili
32	56.5	33.2	358	8	Q9THY2	Q9thy2 gilia latif
33	56.5	33.2	358	8	Q9THY1	Q9thy1 gilia rigid
34	56.5	33.2	358	8	Q9THY0	Q9thy0 gilia scabr
35	56.5	33.2	358	8	Q9THX9	Q9thx9 gymnosteris
36	56.5	33.2	358	8	Q9THX8	Q9thx8 loeselia gl
37	56.5	33.2	358	8	Q8SMK4	Q8smk4 gilia subnu
38	56.5	33.2	359	8	Q8SMK8	Q8smk8 eriastrum w
39	56.5	33.2	360	8	Q8SMK0	Q8smk0 leptodactyl
40	56.5	33.2	360	8	Q8SMJ5	Q8smj5 linanthus p
41	55.5	32.6	358	8	Q32834	Q32834 phlox hoodi
42	55.5	32.6	358	8	Q8SMJ4	Q8smj4 loeseliast
43	55	32.4	358	8	Q32266	Q32266 gilia conge
44	54.5	32.1	358	8	Q8SMK5	Q8smk5 gilia luden
45	54	31.8	328	16	Q92CJ2	Q92cj2 listeria in

ALIGNMENTS

RESULT 1

Q91X13 ID Q91X13 PRELIMINARY; PRT; 110 AA.
AC Q91X13;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Insulin.
OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Scurinae;
OC Spermophilus.
OX NCBI_TaxID=43179;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Tredrea M.M., Buck M.J., Guhaniyogi J., Squire T.L., Andrews M.T.;
RT 'Regulation of PDK4 expression in a hibernating mammal.';
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL: AY038604; AAK72558.1;
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin; 1.
DR PROSITE: PS00262; INSULIN; UNKNOWN_1.
SQ SEQUENCE 110 AA; 12004 MW; 4511768D6622BEE5 CRC64;

Query Match 55.9%; Score 95; DB 11; Length 110;

Best Local Similarity 65.5%; Pred. No. 5.7e-06; Indels 0; Gaps 0;
Matches 19; Conservative 3; Mismatches 7;

Oy 2 VNFHLYGSHIREWLYLVAGERGFNFDPKT 30

Db 26 VNHLCGSHLVREALYLCGERGFFTPKS 54

RESULT 2

Q8WNW6 ID Q8WNW6 PRELIMINARY; PRT; 110 AA.

Q8WNW6;
AC 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Preproinsulin.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PANCREAS.
RA Okamoto S., Morimatsu M.;
RT "cat insulin".
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB043535; BAB84110.1;
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; UNKNOWN 1.
SQ SEQUENCE 110 AA; 12069 MW; 95FB6E170C7BECA4 CRC64;

Query Match 55.3%; Score 94; DB 6; Length 110;
Best Local Similarity 67.9%; Pred. No. 7.9e-06;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPK 29
II III: I IIII IIIII: II
DB 26 VNOHLCGSHLVEALFLVCGRGFFFTPK 53

RESULT 3
Q98TB0
ID Q98TB0 PRELIMINARY; PRT; 111 AA.
AC Q98TB0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Preproinsulin (Fragment).
OS Chitala chitala (clown knifefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Moptopteridae; Chitala.
OX NCBI_TaxID=112163;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21203577; PubMed-11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid.";
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF199586; AAK28710.1;
DR HSSP; P01308; ILPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; IIGF; 1.
FT NON_TER 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match 54.18%; Score 92; DB 13; Length 111;
Best Local Similarity 66.7%; Pred. No. 1.6e-05;
Matches 18; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 NFHLYGSHIREWLYLVAGRGFNFDPK 29
I IIII: I IIII IIIII: II
DB 26 NQHLCGSHLVEALFLVCGRGFFFTPK 52

RESULT 4
Q98TA9
ID Q98TA9 PRELIMINARY; PRT; 87 AA.
AC Q98TA9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Preproinsulin (Fragment).
OS Gnathonemus petersii.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Mormyridae; Gnathonemus.
OX NCBI_TaxID=42645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21203577; PubMed-11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid.";
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF199587; AAK28711.1;
DR HSSP; P01308; IHIS.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
FT NON_TER 1
FT NON_TER 87
SQ SEQUENCE 87 AA; 9874 MW; FF448ED35D2453F5 CRC64;

Query Match 51.2%; Score 87; DB 13; Length 87;
Best Local Similarity 65.4%; Pred. No. 6.2e-05;
Matches 17; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 5 HLYGSHIREWLYLVAGRGFNFDPK 30
II III: I III IIIII: I I
DB 5 HLCGSHLVEALFLVCGRGFFFTPK 30

RESULT 5
Q98TA8
ID Q98TA8 PRELIMINARY; PRT; 110 AA.
AC Q98TA8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Preproinsulin.
OS Pantodon buchholzi (Butterflyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Pantodontidae; Pantodon.
OX NCBI_TaxID=8276;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21203577; PubMed-11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid.";
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF199588; AAK28712.1;
DR HSSP; P01308; IHIS.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
SQ SEQUENCE 110 AA; 12324 MW; BDECCD659D872E06 CRC64;

RT-MOLECULAR

RT osteoglossomorphs and a cyprinid.";
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF199585; AAK28709.1; -.
DR HSP; P01308; ILPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF000049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
FT NON_TER 108 108
SQ SEQUENCE 108 AA; 11873 MW; E426310696FBAFC8 CRC64;

Query Match 43.5%; Score 74; DB 13; Length 108;
Best Local Similarity 56.0%; Pred. No. 0.0056;
Matches 14; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 HLYGSHIREWLYVAGERGFNDPK 29
II III: : IIII I II : : II
DB 28 HLCGSHLDALYVCGPTGFFYNPK 52

RESULT 10

Q9DDE5 PRELIMINARY; PRT; 108 AA.
AC Q9DDE5
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Insulin precursor.
GN INS.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99425190; PubMed-10495291;
RA Argenton F., Zecchin E., Bortolussi M.;
RT "Early appearance of pancreatic hormone-expressing cells in the
RL zebrafish embryo.";
RL Mech. Dev. 87:217-221(1999).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AJ237750; CAC20109.1; -.
DR HSP; P01308; ILPH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF000049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Signal.

FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 53 INSULIN B CHAIN.
FT CHAIN 86 108 INSULIN A CHAIN.
SQ SEQUENCE 108 AA; 11904 MW; 3195289E72AD6D25 CRC64;

Query Match 43.5%; Score 74; DB 13; Length 108;
Best Local Similarity 56.0%; Pred. No. 0.0056;
Matches 14; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 HLYGSHIREWLYVAGERGFNDPK 29
II III: : IIII I II : : II
DB 28 HLCGSHLDALYVCGPTGFFYNPK 52

RESULT 11

Q902N4 PRELIMINARY; PRT; 108 AA.
ID Q902N4
AC Q902N4
DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Preproinsulin.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Catla.
OX NCBI_TaxID=72446;
RN [1]
RP SEQUENCE FROM N.A.
RA Bhattacharya S., Roy S.S., Dasgupta S., Ravikumar L., Mukherjee M.,
RA Bandyopadhyaya I., Wakabayashi K.;
RT "A new cell secreting insulin.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF373021; AAK51558.1; -.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF000049; Insulin; 1.
DR PROSITE; PS00262; INSULIN; UNKNOWN_1.
DR PROSITE; PS00262; INSULIN; 11881 MW; D713026E22EF5D59 CRC64;
SQ SEQUENCE 108 AA; 11881 MW; D713026E22EF5D59 CRC64;

Query Match 43.5%; Score 74; DB 13; Length 108;
Best Local Similarity 56.0%; Pred. No. 0.0056;
Matches 14; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 HLYGSHIREWLYVAGERGFNDPK 29
II III: : IIII I II : : II
DB 28 HLCGSHLDALYVCGPTGFFYNPK 52

RESULT 12

Q987A7 PRELIMINARY; PRT; 111 AA.
ID Q987A7
AC Q987A7
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Preproinsulin (Fragment).
OS Osteoglossum bicirrhosum (silver arowana).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Osteoglossidae; Osteoglossum.
OX NCBI_TaxID=109271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21203577; PubMed-11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid.";
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF199589; AAK28713.1; -.
DR HSP; P01315; LMPJ.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF000049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12491 MW; AC9E19D2D4866D20 CRC64;

Query Match 42.9%; Score 73; DB 13; Length 111;
Best Local Similarity 52.0%; Pred. No. 0.0081;
Matches 13; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 6 LYGSHIREWLYVAGERGFNDPK 30
I III: : IIII I II : : II
DB 30 LCGSHLDALYVCGDRGFFYSPKS 54

RESULT 13

Q9TCV1
ID AC Q9TCV1 PRELIMINARY; PRT; 151 AA.
RT
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).
GN COL.
OS Metagonimus 'Miyata type'.
OG Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Opisthorchiida; Opisthorchiata; Opisthorchioidea; Heterophyidae;
OC Metagonimus.
OX NCBI_TaxID=87632;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee S.U., Park H.Y., Chung J.Y., Huh S.;
RT "Phylogenetic data of some helminthes by sequence analysis of PCR product of specific genes";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3 AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 4 FERROCYTOCHROME C + 2 H(2)O.
CC -1- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AF096232; AAD54522.1; -
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCIXDASBI.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT NON_TER 1 1
FT NON_TER 151 151
SQ SEQUENCE 151 AA; 16731 MW; 4321828B042B66F0 CRC64;
Query Match 34.7%; Score 59; DB 8; Length 151;
Best Local Similarity 47.6%; Pred. No. 1.1;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY 8 GSHIRENLYLVAGERGFNDP 28
DB 81 GIKVSSWLYMLAGNRGFWD 101
RESULT 14
Q8SMK3
ID AC Q8SMK3 PRELIMINARY; PRT; 328 AA.
RT
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Maturase (Fragment).
GN MATK.
OS Ipomopsis minutiflora.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Polemoniaceae; Ipomopsis.
OX NCBI_TaxID=160044;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson L.A., Schultz J.L., Soltis D.E., Soltis P.S.;
RT "Monophyly and generic relationships of Polemoniaceae based on matk

sequences.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; L48584; AAL77540.1; -
KW Chloroplast. 328 328
FT NON_TER 328 328
SQ SEQUENCE 328 AA; 40104 MW; 7D1503B1F4270D1C CRC64;
Query Match 34.4%; Score 58.5; DB 8; Length 328;
Best Local Similarity 37.5%; Pred. No. 3.1;
Matches 12; Conservative 6; Mismatches 5; Indels 9; Gaps 1;
QY 3 NFELYGSH-----IREWLYLVAGERGFN 25
DB 9 NFELYGSHQHYFYVYPLIFQYIYVLAHGRGLN 40
RESULT 15
Q32253
ID AC Q32253 PRELIMINARY; PRT; 358 AA.
RT
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K) (Fragment).
GN MATK.
OS Gilia australis.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Polemoniaceae; Gilia.
OX NCBI_TaxID=40768;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson L.A., Soltis D.E.;
RT "Phylogenetic inference using matk sequences.";
RL Ann. Mo. Bot. Gard. 0:0-0(1994).
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II INTRONS (BY SIMILARITY).
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS, AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY MITOCHONDRIAL INTRONS.
DR EMBL; L34178; AAA84271.1; -
DR InterPro; IPR002866; MATK_N.
DR Pfam; PF01824; MATK_N; 1.
KW Chloroplast; mRNA processing.
FT NON_TER 358 358
SQ SEQUENCE 358 AA; 43172 MW; F17764E99E359FE3 CRC64;
Query Match 34.4%; Score 58.5; DB 8; Length 358;
Best Local Similarity 37.5%; Pred. No. 3.4;
Matches 12; Conservative 5; Mismatches 6; Indels 9; Gaps 1;
QY 3 NFELYGSH-----IREWLYLVAGERGFN 25
DB 7 NFELYGSHQHYFYVYPLIFQYIYVLAHGRGLN 38
RESULT 16
Q32268
ID AC Q32268 PRELIMINARY; PRT; 358 AA.
RT
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K) (Fragment).
GN MATK.
OS Gilia caruifolia.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Polemoniaceae; Gilia.
OX NCBI_TaxID=40772;

```

RN SEQUENCE FROM N.A.
RP TISSUE=LEAF;
RA Johnson L.A., Soltis D.E.;
RT "Phylogenetic inference using matk sequences.";
RL Ann. Mo. Bot. Gard. 0:0-0(1994).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; L34183; AA84274.1; -.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
FT NON_TER 358
SQ SEQUENCE 358 AA; 43200 MW; D230E06B2532C95F CRC64;

Query Match 34.4%; Score 58.5; DB 8; Length 358;
Best Local Similarity 37.5%; Pred. No. 3.4;
Matches 12; Conservative 5; Mismatches 6; Indels 9; Gaps 1;

QY 3 NFELYGSH-----IRENLYVAGERFN 25
DB 7 NFELYGSHQHYFYVPLIFQEIYIAIADHRLN 38

RESULT 17
Q32356 PRELIMINARY; PRT; 358 AA.
AC Q32356;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Probable intron maturase (Maturase K) (Fragment).
GN MATK.
OS Gallia splendens.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Polemoniaceae; Gallia.
OX NCBI_TaxID=40779;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson L.A., Soltis D.E.;
RT "Phylogenetic inference using matk sequences.";
RL Ann. Mo. Bot. Gard. 0:0-0(1994).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; L34191; AA84280.1; -.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
FT NON_TER 358
SQ SEQUENCE 358 AA; 43188 MW; 3D721EC9AF100CBB CRC64;

Query Match 34.4%; Score 58.5; DB 8; Length 358;
Best Local Similarity 37.5%; Pred. No. 3.4;
Matches 12; Conservative 5; Mismatches 6; Indels 9; Gaps 1;

QY 3 NFELYGSH-----IRENLYVAGERFN 25
DB 7 NFELYGSHQHYFYVPLIFQEIYIAIADHRLN 38

RESULT 18
Q32809 PRELIMINARY; PRT; 360 AA.
AC Q32809;

```

```

DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K) (Fragment).
GN MATK.
OS Polemonium californicum (California Jacob's-ladder).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Polemoniaceae; Polemonium.
OX NCBI_TaxID=40738;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson L.A., Soltis D.E.;
RT "Phylogenetic inference using matk sequences.";
RL Ann. Mo. Bot. Gard. 0:0-0(1994).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; L34204; AA84557.1; -.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
FT NON_TER 360
SQ SEQUENCE 360 AA; 43607 MW; FADF8EC5D5D8D636 CRC64;

Query Match 33.8%; Score 57.5; DB 8; Length 360;
Best Local Similarity 40.6%; Pred. No. 4.7;
Matches 13; Conservative 4; Mismatches 6; Indels 9; Gaps 1;

QY 3 NFELYGSHIR-----EWLYVAGERFN 25
DB 7 NFELYGSHQHYFYVPLIFQEIYIAIADHRLN 38

RESULT 19
Q3THX7 PRELIMINARY; PRT; 360 AA.
AC Q3THX7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K) (Fragment).
GN MATK.
OS Polemonium pauciflorum.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Polemoniaceae; Polemonium.
OX NCBI_TaxID=64434;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson L.A., Soltis D.E., Soltis P.S.;
RT "Monophyly and generic relationships of Polemoniaceae based on matk
sequences.";
RL Am. J. Bot. 83:1207-1224(1996).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; L48602; AAF22870.1; -.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
FT NON_TER 360
SQ SEQUENCE 360 AA; 43554 MW; AEFBE38B4867B2D8 CRC64;

Query Match 33.8%; Score 57.5; DB 8; Length 360;

```

Best Local Similarity 40.6%; Pred. No. 4.7;
Matches 13; Conservative 4; Mismatches 6; Indels 9; Gaps 1;

QY 3 NFHLGSHIR-----EWLYLVAGERGFN 25
Db 7 NFELIGSOORFVYPLIFQEIYALAHDRGLN 38

RESULT 20
Q9TCV2
ID Q9TCV2 PRELIMINARY; PRT; 148 AA.
AC Q9TCV2
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TremBLrel. 20, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).
GN COI.
OS Metagonimus yokogawai.
OG Mitochondrion.
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC Opisthorchiida; Opisthorchiata; Opisthorchioidea; Heterophyidae;
OC Metagonimus.
OX NCBI_TaxID=84529;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee S.U., Park H.Y., Chung J.Y., Huh S.;
RT "Phylogenetic data of some helminthes by sequence analysis of PCR
product of specific genes.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
AND COPPER B (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) - 4 FERROCYTOCHROME
C + 2 H(2)O.
CC -1- COFACTOR: HEMES A, A3, AND COPPER B (BY SIMILARITY).
CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AF096230; AA054521.1; -
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COX1; 1
DR PRINTS; PR01165; CYCOXIDASE1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT NON_TER 1 148
FT NON_TER 148 148
SQ SEQUENCE 148 AA; 16196 MW; EAB3E3828374F024 CRC64;

Query Match 33.5%; Score 57; DB 8; Length 148;
Best Local Similarity 47.6%; Pred. No. 2.1;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 8 GSHREWLYLVAGERGFNFD 28
Db 82 GIKVFSWLYLVAGSRGWDP 102

RESULT 21
Q8TZA8
ID Q8TZA8 PRELIMINARY; PRT; 154 AA.
AC Q8TZA8
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Ribosomal protein L15.
GN RPLO OR MK0027.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;

Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-AV19 / DSM 6324 / JCM 9639;
RX MEDLINE-21927647; PubMed-11930014;
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natile D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AE010303; AA01244.1; -
KW Complete proteome.
SQ SEQUENCE 154 AA; 17040 MW; 24364D7588EE902A CRC64;

Query Match 33.5%; Score 57; DB 17; Length 154;
Best Local Similarity 39.3%; Pred. No. 2.2;
Matches 11; Conservative 5; Mismatches 6; Indels 6; Gaps 1;

QY 8 GSHREWLYLV-----GERGFNFD 29
Db 40 GSHKHKWFVIRKYPDPDFGKGFGRNRP 67

RESULT 22
Q9KBM0
ID Q9KBM0 PRELIMINARY; PRT; 299 AA.
AC Q9KBM0
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Transcriptional regulator (Arac/XylS family).
GN BH1906.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-C-125 / JCM 9153;
RX MEDLINE-20512582; PubMed-11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- SIMILARITY: BELONGS TO THE ARAC/XYLIS FAMILY OF TRANSCRIPTIONAL
REGULATORS.
DR EMBL; AP001513; BAB05625.1; -
DR InterPro; IPR000005; HTHARAC.
DR Pfam; PF00165; HTH_Arac; 2.
DR PRINTS; PR00032; HTHARAC.
DR SMART; SM00342; HTH_Arac; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 2.
DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
KW DNA-binding; Transcription regulation; Complete proteome.
SQ SEQUENCE 299 AA; 35029 MW; 4D9FE2FB916F4884 CRC64;

Query Match 33.5%; Score 57; DB 16; Length 299;
Best Local Similarity 39.3%; Pred. No. 4.5;
Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNFD 29
Db 31 VILCHWHHEWITWEAGEAEFQIGTK 58

RESULT 23
Q32181
ID Q32181 PRELIMINARY; PRT; 358 AA.

AC Q32181;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K) (Fragment).
GN MAYK.
OS Eriastrum densifolium subsp. mohavense.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Polemoniaceae; Eriastrum.
OX NCBI_TaxID=40764;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Johnson L.A., Soltis D.E.; Soltis D.E.;
RT "Phylogenetic inference using matk sequences.";
RL Ann. Mo. Bot. Gard. 0:0-0(1994).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL: L34184; AAA84244.1; -
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
FT NON_TER 358
SQ SEQUENCE 358 AA; 43549 MW; D6B8A9726CE8F04F CRC64;
Query Match 33.2%; Score 56.5; DB 8; Length 358;
Best Local Similarity 37.5%; Pred. No. 6.5;
Matches 12; Conservative 5; Mismatches 6; Indels 9; Gaps 1;
QY 3 NFHLYGSH-----IREWLYVAGERGN 25
DB 7 NFELYGSOQHFFVYPLIFQEIYIALAHDRLN 38
RESULT 24
Q32287 PRELIMINARY; PRT; 358 AA.
AC Q32287;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K) (Fragment).
GN MAYK.
OS Gilla filiformis.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Polemoniaceae; Gilla.
OX NCBI_TaxID=40773;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Johnson L.A., Soltis D.E.;
RT "Phylogenetic inference using matk sequences.";
RL Ann. Mo. Bot. Gard. 0:0-0(1994).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL: L34185; AAA84275.1; -
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
FT NON_TER 358
SQ SEQUENCE 358 AA; 43393 MW; 31F4E2CAD3B17CE2 CRC64;
Query Match 33.2%; Score 56.5; DB 8; Length 358;

Best Local Similarity 37.5%; Pred. No. 6.5;
Matches 12; Conservative 5; Mismatches 6; Indels 9; Gaps 1;
QY 3 NFHLYGSH-----IREWLYVAGERGN 25
DB 7 NFELYGSOQHFFVYPLIFQEIYIALAHDRLN 38
RESULT 25
Q32290 PRELIMINARY; PRT; 358 AA.
AC Q32290;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K) (Fragment).
GN MAYK.
OS Gilla hutchinsifolia.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Ericales; Polemoniaceae; Gilla.
OX NCBI_TaxID=40774;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Johnson L.A., Soltis D.E.;
RT "Phylogenetic inference using matk sequences.";
RL Ann. Mo. Bot. Gard. 0:0-0(1994).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL: L34186; AAA84276.1; -
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
FT NON_TER 358
SQ SEQUENCE 358 AA; 43473 MW; 47E620D3D4F76966 CRC64;
Query Match 33.2%; Score 56.5; DB 8; Length 358;
Best Local Similarity 37.5%; Pred. No. 6.5;
Matches 12; Conservative 5; Mismatches 6; Indels 9; Gaps 1;
QY 3 NFHLYGSH-----IREWLYVAGERGN 25
DB 7 NFELYGSOQHFFVYPLIFQEIYIALAHDRLN 38
Search completed: December 23, 2002, 07:15:43
Job time : 31.4118 secs

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 14:29:57 ; Search time 30.02 Seconds
(without alignments)
188.700 Million cell updates/sec

Title: US-09-574-443-7

Perfect score: 275

Sequence: 1 NLVEQASTQASLYQIYNFD.....IREWLYLVAGERGDFDKPT 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_032802.*
- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
 - 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
 - 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
 - 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
 - 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
 - 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
 - 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
 - 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
 - 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
 - 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
 - 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
 - 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
 - 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
 - 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
 - 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
 - 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
 - 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
 - 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
 - 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
 - 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
 - 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
 - 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	170	61.8	30	22	AA198851
2	151	54.9	30	22	AA198874
3	149	54.2	51	14	AA198049
4	149	54.2	51	22	AA1991205
5	149	54.2	61	7	AA1960154
6	149	54.2	78	7	AA1960153
7	149	54.2	84	7	AA1960248
8	144	52.4	51	14	AA198048
9	140	50.9	51	14	AA198050
10	132	48.0	30	22	AA198876
11	128	46.5	51	14	AA198852

12	126	45.8	51	14	AA198851	Phosphorylated rat
13	121	44.0	30	22	AA198872	Mutant human insul
14	115.5	42.0	86	13	AA192514	ACB-Proinsulin, S
15	114	41.5	30	22	AA198859	Mutant human insul
16	113	41.1	30	22	AA198867	Mutant human insul
17	110	40.0	30	22	AA198853	Mutant human insul
18	110	40.0	30	22	AA198857	Mutant human insul
19	110	40.0	117	21	AA1969788	MWPSp-MWPMp10-Met-
20	106.5	38.7	124	21	AA197081	Insulin A-B chain
21	106.5	38.7	124	22	AA1965634	MWPSp-MWPMp9-GSLQP
22	105	38.2	21	22	AA198850	Mutant human insul
23	104	37.8	30	17	AA1988748	Asp(1) human insul
24	104	37.8	30	22	AA198868	Mutant human insul
25	104	37.8	30	22	AA198869	Mutant human insul
26	103	37.5	30	22	AA198866	Mutant human insul
27	101	36.7	96	15	AA1968899	Human pro-insulin
28	101	36.7	96	16	AA1978662	Fusion protein con
29	99.5	36.2	125	18	AA192422	EEAEPK-M15 insul
30	99	36.0	29	3	AA1920139	Sequence of des-ph
31	99	36.0	29	6	AA1950834	Sequence of des-ph
32	99	36.0	30	6	AA1950826	Sequence of human
33	99	36.0	30	7	AA1960909	Derivative of insu
34	99	36.0	30	7	AA1961335	Modified insulin B
35	99	36.0	30	11	AA1907755	Insulin analogue B
36	99	36.0	30	11	AA1904040	Amino acids 1-30 o
37	99	36.0	30	13	AA1930650	Phosphorylated hum
38	99	36.0	30	13	AA1926470	Sequence of human
39	99	36.0	30	14	AA1944454	Human insulin B-ch
40	99	36.0	30	15	AA1962642	Bovine and human i
41	99	36.0	30	15	AA1963881	Human insulin B-ch
42	99	36.0	30	15	AA1968896	Human insulin B-ch
43	99	36.0	30	17	AA1960000	Asymmetry sequence
44	99	36.0	30	17	AA1960002	Asymmetry sequence
45	99	36.0	30	19	AA1963755	Human insulin prot

ALIGNMENTS

RESULT 1	
AA198851	AA198851 standard; peptide; 30 AA.
XX	AA198851;
XX	AC
XX	16-MAR-2001 (first entry)
DT	
XX	
DE	Mutant human insulin (IA protein) cysl B chain.
XX	
KW	Human insulin; insulin activity protein; IA protein;
KW	diabetes mellitus; type 1; type 2; stability;
KW	disulphide bond; cysteine replacement; mutant; muten.
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
PN	WO200069901-A2.
XX	
PD	23-NOV-2000.
XX	
PF	19-MAY-2000; 2000WO-US13764.
XX	
PR	19-MAY-1999; 99US-0134930.
XX	
FA	(XENC-) XENCOR INC.
XX	
PI	Anti-diabetic pept
XX	Insulin-like anti-
DR	Human proinsulin.
XX	
PT	Non-naturally occurring protein with insulin activity useful for
PT	treating type 1 and type 2 diabetes, comprising amino acid
PT	substitutions as compared to native human insulin and having enhanced

PT stability -
 XX Claim 10; Fig 3A; 95pp; English.
 PS
 CC The invention relates to novel non-naturally occurring mature human
 CC insulin mutants, designated insulin activity (IA) proteins in the
 CC specification (AAB48850-B48876), which have altered properties (e.g.,
 CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
 CC altered ability to oligomerise) compared to wild-type mature human
 CC insulin (AAB48848, AAB48849) but which are still capable of binding to
 CC an insulin receptor. The insulin mutants of the invention have less than
 CC 98% sequence identity to wild-type human insulin, but have a 3D
 CC structure which substantially corresponds to that of wild-type human
 CC insulin. The invention also relates to recombinant nucleic acids encoding
 CC the insulin mutants, expression vectors and host cells comprising mutant
 CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
 CC and a pharmaceutical composition comprising an insulin mutant of the
 CC invention. The mature human insulin mutants are useful for treating
 CC insulin-responsive conditions and disorders of carbohydrate metabolism
 CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
 CC insulin mutants are used in gene therapy techniques for treating these
 CC conditions. The insulin mutants are more stable compared to wild-type
 CC mature human insulin, or have an improved ability to form insulin
 CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
 CC insulin mutants in which one or more cysteine residue has been
 CC replaced, thus preventing the formation of at least one disulphide bond
 CC and improving stability and activity.
 XX Sequence 30 AA;
 SQ

Query Match 61.8%; Score 170; DB 22; Length 30;
 Best Local Similarity 100.0%; Pred. No. 1.5e-14;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 22 DVNFHLYGSHIREWLYLVAGERGFNDPKT 51
 Db 1 dvnfhlygshirewlylvagernfndpkt 30

RESULT 2
 AAB48874
 ID AAB48874 standard; peptide; 30 AA.
 XX
 AC AAB48874;
 XX
 XX 16-MAR-2001 (first entry)
 XX Mutant human insulin (IA protein) trz_07 B chain.
 XX Human insulin; insulin activity protein; IA protein;
 KW diabetes mellitus; type 1; type 2; stability;
 KW mutant; mutein.
 XX Homo sapiens.
 OS Synthetic.
 PN WO200069901-A2.
 XX
 XX 23-NOV-2000.
 PD 19-MAY-2000; 2000WO-US13764.
 XX 19-MAY-1999; 990US-0134930.
 PR (XENC-) XENCOR INC.
 XX Dahiyyat BI;
 PI WPI; 2001-025004/03.
 XX Non-naturally occurring protein with insulin activity useful for
 PT treating type 1 and type 2 diabetes, comprising amino acid
 PT

PT substitutions as compared to native human insulin and having enhanced
 PT stability -
 XX Claim 10; Fig 5B; 95pp; English.
 PS
 CC The invention relates to novel non-naturally occurring mature human
 CC insulin mutants, designated insulin activity (IA) proteins in the
 CC specification (AAB48850-B48876), which have altered properties (e.g.,
 CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
 CC altered ability to oligomerise) compared to wild-type mature human
 CC insulin (AAB48848, AAB48849) but which are still capable of binding to
 CC an insulin receptor. The insulin mutants of the invention have less than
 CC 98% sequence identity to wild-type human insulin, but have a 3D
 CC structure which substantially corresponds to that of wild-type human
 CC insulin. The invention also relates to recombinant nucleic acids encoding
 CC the insulin mutants, expression vectors and host cells comprising mutant
 CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
 CC and a pharmaceutical composition comprising an insulin mutant of the
 CC invention. The mature human insulin mutants are useful for treating
 CC insulin-responsive conditions and disorders of carbohydrate metabolism
 CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
 CC insulin mutants are used in gene therapy techniques for treating these
 CC conditions. The insulin mutants are more stable compared to wild-type
 CC mature human insulin, or have an improved ability to form insulin
 CC hexamers. Sequences AAB48871-B48876 represent the A and B chains of
 CC insulin mutants which have improved stability.
 XX Sequence 30 AA;
 SQ

Query Match 54.9%; Score 151; DB 22; Length 30;
 Best Local Similarity 90.0%; Pred. No. 3.6e-12;
 Matches 27; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 22 DVNFHLYGSHIREWLYLVAGERGFNDPKT 51
 Db 1 dknfhlygshirewlylvagernfndpkt 30

RESULT 3
 AAR40849
 ID AAR40849 standard; peptide; 51 AA.
 XX
 AC AAR40849;
 XX
 XX 07-MAR-1994 (first entry)
 DT Phosphorylated human insulin.
 DE Diabetes mellitus; treatment; hypoglycaemia; hyperglycaemia;
 XX reduction; induction.
 KW Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..21 /note= "A chain"
 FT Peptide 22..51 /note= "B chain"
 FT Disulfide-bond 6..11 /note= "Cys-Cys cross links"
 FT Disulfide-bond 7..28 /note= "Cys-Cys cross links"
 FT Disulfide-bond 20..40 /note= "Cys-Cys cross links"
 XX US5242900-A.
 XX 07-SEP-1993.
 PD 30-MAY-1991; 91US-0707542.
 XX 30-MAY-1991; 91US-0707542.
 PR

XX (ALBI/) ALBISSER AM.
 XX Albiisser AM;
 XX WPI; 1993-295253/37.
 XX Phosphorylated insulin for treating diabetes mellitus - reduces
 PT hyperglycaemia without inducing hypoglycaemia
 XX Claim 1; Fig 1; 13pp; English.
 XX The sequence is that of phosphorylated human insulin which
 CC may be used to treat diabetes mellitus without inducing hypoglycaemia.
 CC The insulin is given by subcutaneous injection. One or more of the
 CC serine and/or threonine residues are phosphorylated as O-phosphates.
 XX Sequence 51 AA;
 SQ

Query Match 54.2%; Score 149; DB 14; Length 51;
 Best Local Similarity 62.0%; Pred. No. 1.2e-11;
 Matches 31; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 2 LVFQASTSQASLYQIYNFNDVNFHLYGSHIREWLYLVAGERGFNFDPKT 51
 Db :||| || ||||| : || || ||||| : ||||| ||||| : |||||
 2 iveqcetsiclsyqlencfncvnhlcgshivealylvcgergfftyptkt 51

RESULT 4
 AAB91205
 ID AAB91205 standard; Peptide; 51 AA.
 XX AC AAB91205;
 XX 22-JUN-2001 (first entry)
 XX Insulin and insulin-like peptide SEQ ID NO:379.
 XX Protection; endogenous therapeutic peptide; peptidase; conjugation;
 KW blood component; modification; succinimidy; maleimido group; amino;
 KW hydroxyl; thiol; hormone; growth factor; neurotransmitter.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200069900-A2.
 XX 23-NOV-2000.
 XX 17-MAY-2000; 2000WO-US13576.
 XX 17-MAY-1999; 99US-0134406.
 PR 10-SEP-1999; 99US-0153406.
 PR 15-OCT-1999; 99US-0159783.
 XX (CONJ-) CONJUCHEM INC.
 XX Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudeau K;
 XX WPI; 2001-112059/12.
 XX Modifying and attaching therapeutic peptides to albumin prevents
 PT peptidase degradation, useful for increasing length of in vivo activity
 PT
 XX Disclosure; Page 321; 733pp; English.
 XX The present invention describes a modified therapeutic peptide (I)
 CC comprising a therapeutically active amino acid region (III) and a
 CC reactive group (II) (e.g. succinimidy and maleimido groups) attached to
 CC a less therapeutically active amino acid region (IV), which covalently
 CC bonds with amino/hydroxyl/thiol groups on blood components to form a

CC peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
 CC (I) are useful for modifying therapeutic peptides e.g. hormones, growth
 CC factors and neurotransmitters, to protect them from peptidase activity
 CC in vivo for the treatment of various disorders. Endogenous therapeutic
 CC peptides are not suitable as drug candidates as they require frequent
 CC administration due to rapid degradation by peptidases in the body.
 CC Modifying and attaching therapeutic peptides to albumin prevents or
 CC reduces the action of peptidases to increase length of activity (half
 CC life) and specificity as bonding to large molecules decreases
 CC intracellular uptake and interference with physiological processes.
 CC AAB90829 to AAB92441 represent peptides which can be used in the
 CC exemplification of the present invention.
 XX Sequence 51 AA;
 SQ

Query Match 54.2%; Score 149; DB 22; Length 51;
 Best Local Similarity 62.0%; Pred. No. 1.2e-11;
 Matches 31; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 2 LVFQASTSQASLYQIYNFNDVNFHLYGSHIREWLYLVAGERGFNFDPKT 51
 Db :||| || ||||| : || || ||||| : ||||| ||||| : |||||
 2 iveqcetsiclsyqlencfncvnhlcgshivealylvcgergfftyptkt 51

RESULT 5
 AAP60154
 ID AAP60154 standard; peptide; 61 AA.
 XX AC AAP60154;
 XX 01-JUL-1991 (first entry)
 XX Anti-diabetic peptide.
 XX Diabetes; insulin-like peptide; proinsulin;
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH Disulfide-bond 15..20
 FT Disulfide-bond 29..49
 FT Disulfide-bond 16..37
 XX EP171887-A.
 XX 19-FEB-1986.
 XX 12-JUN-1985; 85EP-0304165.
 XX 14-JUN-1984; 84US-0620781.
 XX (ELIL) ELI LILLY & CO.
 XX Frank BH, Pekar AH;
 XX WPI; 1986-049748/08.
 XX New anti-diabetic peptide cpds. - prepd. from human pro-insulin
 PT by treatment with e.g. chymotrypsin
 XX Claim 1; page 16; 19pp; English.
 XX The peptide has insulin-like anti-diabetic activity which is greater
 CC than that of known proinsulin. X is RR (A) or RREADQGVGVGGGAGS-
 CC LQPL (B) or is absent. The peptide is produced by: treating human
 CC proinsulin with human chymotrypsin to produce the peptide in which X
 CC is (B); by treating human proinsulin with trypsin to produce (32-33
 CC split) human proinsulin which is treated with chymotrypsin to yield the
 CC peptide in which X is (A); or by treating the peptide in which X is (A)
 CC with carboxypeptidase B to produce the peptide in which X is absent.
 XX Sequence 61 AA;

Query Match 54.2%; Score 149; DB 7; Length 61;
 Best Local Similarity 62.0%; Pred. No. 1.4e-11;
 Matches 31; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDNDVNFHLYGSHIREWLYLVAGERGFNFDPKT 51
 :||| || ||||: | || || ||||: | |||| ||||: |||
 Db 11 iveqctscslyqlenycnfvnqhlcgshlvealyivcgergfftyptk 60

RESULT 6

AAP60153
 ID AAP60153 standard; peptide; 78 AA.

XX AC AAP60153;

XX PT '01-JUL-1991 (first entry)

Insulin-like anti-diabetic peptide.

KW Diabetes; insulin-like peptide; human proinsulin;

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT Disulfide-bond 6..11

FT Disulfide-bond 7..28

FT Disulfide bond 20..40

XX EP171886-A.

XX PD 19-FEB-1986.

XX PF 12-JUN-1985; 85EP-0304162.

XX PR 14-JUN-1984; 84US-0620780.

XX PA (ELIL) ELI LILLY & CO.

XX PI Frank BH, Pekar AH;

XX WPI; 1986-049747/08.

XX New insulin-like anti-diabetic peptide cpds. - prepd. by treating
 PT pro-insulin with trypsin and opt. carboxypeptidase B or
 PT chymotrypsin

Claim 1; page 18; 2lpp; English.

XX The peptide and its salts have insulin-like anti-diabetic activity
 CC which is greater than that of human proinsulin. X is absent or is
 CC ALEGSIQ (A) or ALEGSIQKR (B). The peptide is prepd. by: treating human
 CC proinsulin with trypsin to produce the peptide in which X is (B) (1);
 CC treating (1) with carboxypeptidase B to yield the peptide in which X
 CC is (A); or treating (1) with chymotrypsin to produce the peptide in
 CC which X is absent.

XX Sequence 78 AA;

Query Match 54.2%; Score 149; DB 7; Length 78;
 Best Local Similarity 62.0%; Pred. No. 1.9e-11;
 Matches 31; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDNDVNFHLYGSHIREWLYLVAGERGFNFDPKT 51
 :||| || ||||: | || || ||||: | |||| ||||: |||
 Db 2 iveqctscslyqlenycnfvnqhlcgshlvealyivcgergfftyptk 51

RESULT 7

AAP60248
 ID AAP60248 standard; peptide; 84 AA.

XX AAP60248;
 AC 06-JUL-1991 (first entry)
 DT Human proinsulin.
 DE Human proinsulin; diabetes;
 KW Homo sapiens.
 XX Key Location/Qualifiers
 PH Disulfide-bond 39..44
 FT Disulfide-bond 40..61
 FT Disulfide-bond 53..73
 XX EP171147-A.

XX PD 12-FEB-1986.

XX PF 12-JUN-1985; 85EP-0304164.

XX PR 14-JUN-1984; 84US-0620782.

XX PA (ELIL) ELI LILLY & CO.

XX PI Frank BH, Pekar AH;

XX WPI; 1986-043674/07.

XX New antidiabetic peptide(s) - obt'd. from human pro-insulin by
 PT enzymatic treatment

XX Claim 1; page 15; 17pp; English.

XX The peptide is designated (32-33 split) human proinsulin. It has
 CC antidiabetic activity at a level greater than that of natural human
 CC proinsulin. It is obt'd. from the natural molecule by trypsin
 CC treatment. It is useful for the treatment of diabetes by
 CC parenteral routes.

XX Sequence 84 AA;

Query Match 54.2%; Score 149; DB 7; Length 84;
 Best Local Similarity 62.0%; Pred. No. 2.1e-11;
 Matches 31; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDNDVNFHLYGSHIREWLYLVAGERGFNFDPKT 51
 :||| || ||||: | || || ||||: | |||| ||||: |||
 Db 35 iveqctscslyqlenycnfvnqhlcgshlvealyivcgergfftyptk 84

RESULT 8

AAR40848
 ID AAR40848 standard; peptide; 51 AA.

XX AAR40848;

XX DT 07-MAR-1994 (first entry)

XX Phosphorylated pig insulin.

XX Diabetes mellitus; treatment; hypoglycaemia; hyperglycaemia;
 KW reduction; induction; porcine.

XX Sus scrofa domestica.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Peptide /note= "A chain"

FT Peptide 22..51

FT Peptide /note= "B chain"

```

XX PF 30-MAY-1991; 91US-0707542.
XX XX
XX PR 30-MAY-1991; 91US-0707542.
XX XX
XX PA (ALBI/) ALBISSER AM.
XX XX
XX PI Albißer AM;
XX XX
XX DR WPI; 1993-295253/37.
XX XX
XX PT phosphorylated insulin for treating diabetes mellitus - reduces hyperglycaemia without inducing hypoglycaemia
XX XX
XX PS Claim 1; Fig 1; 13pp; English.
XX CC
XX CC The sequence is that of phosphorylated cattle insulin which may be used to treat diabetes mellitus without inducing hypoglycaemia
XX CC The insulin is given by subcutaneous injection. One or more of the serine and/or threonine residues are phosphorylated as O-phosphates.
XX SQ Sequence 51 AA;

Query Match 50.9%; Score 140; DB 14; Length 51;
Best Local Similarity 59.2%; Pred. No. 1.6e-10;
Matches 29; Conservative 5; Mismatches 15; Indels 0; Gaps

QY 2 LVEQASTSQASLYQIYNFDNDVNFHLYGSHIREWLVLAGERGFGNFPDK 50
Db 2 lveqcasvcslyqlenfcfnqdhcshlvealylvcgergfgytpk 50

RESULT 10
AAB48876
ID AAB48876 standard; peptide; 30 AA.
XX AC AAB48876;
XX DT
XX DE 16-MAR-2001 (first entry)
XX KW Mutant human insulin (IA protein) trz_08 B chain.
XX KW Human insulin; Insulin activity protein; IA protein; diabetes mellitus; type 1; type 2; stability; mutant; mutein.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200069901-A2.
XX PD 23-NOV-2000.
XX PF 19-MAY-2000; 2000WO-US13764.
XX PR 19-MAY-1999; 99US-0134930.
XX PA (XENC-) XENCOR INC.
XX PI Dahlyat BI;
XX DR
XX XX WPI; 2001-025004/03.
XX PT Non-naturally occurring protein with insulin activity useful for treating type 1 and type 2 diabetes, comprising amino acid substitutions as compared to native human insulin and having enhanced stability -
XX PS Claim 10; Fig 5C; 95pp; English.
XX CC The invention relates to novel non-naturally occurring mature human insulin mutants, designated insulin activity (IA) proteins in the

```

www

The sequence is that of phosphorylated rat I insulin which may be used to treat diabetes mellitus without inducing hypoglycaemia. The insulin is given by subcutaneous injection. One or more of the serine and/or threonine residues are phosphorylated as O-phosphates.

QY 27 LYGSHIREWLYLVAGERGFDPKT 51
 Db 62 lcgshlvealyvcggrgfftyptk 86

RESULT 15

AA848859
 ID AAB48859 standard; peptide; 30 AA.

XX AC AAB48859;
 XX 16-MAR-2001 (first entry)
 XX DE Mutant human insulin (IA protein) cysd+ B chain.
 XX Human insulin; Insulin activity protein; IA protein;
 KW diabetes mellitus; type 1; type 2; stability;
 KW disulphide bond; cysteine replacement; mutant; muten.

OS Homo sapiens.
 XX Synthetic.

PN WO200069901-A2.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000WO-US13764.

XX 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.

XX Dahiyat BI;

XX WPI; 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for
 PT treating type 1 and type 2 diabetes, comprising amino acid
 PT substitutions as compared to native human insulin and having enhanced
 PT stability -

PS Claim 10; Fig 3E; 95pp; English.

XX The invention relates to novel non-naturally occurring mature human
 CC insulin mutants, designated insulin activity (IA) proteins in the
 CC specification (AAB48850-B48876), which have altered properties (e.g.,
 CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
 CC altered ability to oligomerise) compared to wild-type mature human
 CC insulin (AAB48848, AAB48849) but which are still capable of binding to
 CC an insulin receptor. The insulin mutants of the invention have less than
 CC 98% sequence identity to wild-type human insulin, but have a 3D
 CC structure which substantially corresponds to that of wild-type human
 CC insulin. The invention also relates to recombinant nucleic acids encoding
 CC the insulin mutants, expression vectors and host cells comprising mutant
 CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
 CC and a pharmaceutical composition comprising an insulin mutant of the
 CC invention. The mature human insulin mutants are useful for treating
 CC insulin-responsive conditions and disorders of carbohydrate metabolism
 CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
 CC insulin mutants are used in gene therapy techniques for treating these
 CC conditions. The insulin mutants are more stable compared to wild-type
 CC mature human insulin, or have an improved ability to form insulin
 CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
 CC insulin mutants in which one or more cysteine residue has been
 CC replaced, thus preventing the formation of at least one disulphide bond
 CC and improving stability and activity.

XX Sequence 30 AA;

Query Match 41.5%; Score 114; DB 22; Length 30;
 Best Local Similarity 72.4%; Pred. No. 1.7e-07;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGERGFDPKT 51
 Db 2 vnyhlygshlvealyvcggrgfftyptk 30

Search completed: August 23, 2002, 14:32:04
 Job time: 127 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 14:31:17 ; Search time 18.39 Seconds
(without alignments)
266.479 Million cell updates/sec

Title: US-09-574-443-7
Perfect score: 275
Sequence: 1 NLVEQASTQASLYQIYNFD.....IREWLYLVAGERGFPDPKT 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.5	36.2	96	2	epidermal growth f
2	99	36.0	51	1	insulin - elephant
3	99	36.0	110	1	insulin precursor
4	99	36.0	110	2	insulin precursor
5	99	36.0	110	2	insulin precursor
6	99	36.0	110	2	insulin precursor
7	96	34.9	51	1	insulin - goose
8	96	34.9	81	1	insulin precursor
9	95	34.5	51	1	insulin - hamster
10	95	34.5	103	2	insulin precursor
11	95	34.5	110	1	insulin precursor
12	95	34.5	110	2	insulin precursor
13	94	34.2	51	1	insulin - sperm wh
14	94	34.2	51	1	insulin - finback
15	94	34.2	51	1	insulin - sei whal
16	94	34.2	51	1	insulin - goat
17	94	34.2	51	1	insulin - cat
18	94	34.2	51	2	insulin - North Am
19	94	34.2	51	2	insulin precursor
20	94	34.2	77	1	insulin precursor
21	94	34.2	84	1	insulin precursor
22	94	34.2	86	1	insulin precursor
23	94	34.2	105	1	insulin precursor
24	94	34.2	110	1	insulin precursor
25	93	33.8	51	1	insulin - common s
26	93	33.8	108	2	insulin precursor
27	92	33.5	51	1	insulin - Egyptian
28	91	33.1	51	1	insulin - turkey (
29	91	33.1	51	1	insulin - ostrich

30 91 33.1 51 1 A61129 insulin - black-be
31 91 33.1 51 2 A60414 insulin - slider t
32 91 33.1 107 1 IPCH insulin precursor
33 90 32.7 50 1 INBN2 insulin - skipjack
34 90 32.7 51 1 INBMA insulin - Arabian
35 89 32.4 106 1 IPXL1 insulin I precurs
36 88 32.0 106 1 IPXL2 insulin II precurs
37 85 30.9 52 2 S44469 insulin I1 - North
38 85 30.9 52 2 S44470 insulin I2 - North
39 84 30.5 52 1 INXSA insulin - alligato
40 84 30.5 108 1 INMS1 insulin I precurs
41 84 30.5 110 1 IPRTL insulin I precurs
42 83 30.2 50 1 INFIS insulin - shorthor
43 83 30.2 50 1 INONC insulin [validated
44 83 30.2 50 1 INONC insulin - coho sal
45 83 30.2 51 1 INCB insulin - Chinchil

ALIGNMENTS

RESULT 1

PC7082

epidermal growth factor/single chain insulin fusion protein - Bacillus brevis (frag
C:Species: Bacillus brevis
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C:Accession: PC7082; PC7083
R:Koh, M.; Hanagata, H.; Ebisu, S.; Morihara, K.; Takagi, H.

Biosci. Biotechnol. Biochem. 64, 1079-1081, 2000
A:Title: Use of Bacillus brevis for synthesis and secretion of Des-B30 single-chain
A:Reference number: PC7082; MUID:20335834
A:Accession: PC7082
A:Molecule type: DNA
A:Residues: 1-96 <KOH>
A:Accession: PC7083

A:Molecule type: protein
A:Residues: 19-28 <KO2>
C:Genetics:
A:Gene: egf-sci
C:Superfamily: insulin
C:Keywords: fusion protein

Query Match 36.2%; Score 99.5; DB 2; Length 96;

Best Local Similarity 59.0%; Pred. No. 1.7e-05;

Matches 23; Conservative 3; Mismatches 12; Indels 1; Gaps 1;

Qy 12 SLQIYNFDNDVNFHLYGSHIREWLYLVAGERGFPDPK 50

I:| | | | | | | | | | | | | | | | | | | | | |

Db 38 SMY-TEALDKFVNOHLCGSHLVEALYLVCGERGFFYTPK 75

RESULT 2

INEL

insulin - elephant

C:Species: Elephantidae gen. sp. (elephant)

C:Date: 24-Apr-1984 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999

C:Accession: A01584

R:Smith, L.F.

Am. J. Med. 40, 662-666, 1966

A:Title: Species variation in the amino acid sequence of insulin.

A:Reference number: A90029; MUID:66160119

A:Accession: A01584

A:Molecule type: protein

A:Residues: 1-30;31-51 <SMI>

A:Note: the species of elephant is not given, but it is most probably the Indian el

C:Superfamily: insulin

C:Keywords: hormone; pancreas

F:1-30/Domain: insulin chain B #status experimental <BCH>

F:31-51/Product: insulin #status experimental <MAT>

F:31-51/Domain: insulin chain A #status experimental <ACH>

F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 36.0%; Score 99; DB 1; Length 51;
 Best Local Similarity 69.0%; Pred. No. 9.3e-06;
 Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNFDPKT 51
 II II III: I IIII IIIII: III
 DB 2 VNOHLCGSHLVEALYLVCGRGFFTPKT 30

RESULT 3

IPHU Insulin precursor [validated] - human

N:Alternate names: preproinsulin
 C:Species: Homo sapiens (man)
 C:Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 08-Dec-2000
 C:Accession: A93222; A94253; A93216; A93144; A92075; A91186; I58114; A01579; S58
 ;Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.
 ;ature 284, 26-32, 1980
 ;Title: Sequence of the human insulin gene.
 A:Reference number: A93222; MUID:80120725
 A:Accession: A93222
 A:Molecule type: DNA
 A:Residues: 1-110 <BEL>
 A:Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
 R:Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.
 Science 209, 612-615, 1980
 A:Title: Genetic variation in the human insulin gene.
 A:Reference number: A94253; MUID:80236313
 A:Accession: A94253
 A:Molecule type: DNA
 A:Residues: 1-110 <ULL>
 A:Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
 R:Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.
 Nature 282, 525-527, 1979
 A:Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.
 A:Reference number: A93216; MUID:80054779
 A:Accession: A93216
 A:Molecule type: mRNA
 A:Residues: 1-110 <BEU2>
 A:Cross-references: NID:G186429; PIDN:AAA59172.1; PID:G386828
 R:Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.
 Science 208, 57-59, 1980
 A:Title: Nucleotide sequence of human preproinsulin complementary DNA.
 A:Reference number: A94251; MUID:80147417
 A:Accession: A94251
 A:Molecule type: mRNA
 A:Residues: 1-110 <SUR>
 A:Cross-references: GB:J00265; NID:G186429; PIDN:AAA59172.1; PID:G386828
 R:Nicol, D.S.H.W.; Smith, L.F.
 Nature 187, 483-485, 1960
 A:Title: Amino-acid sequence of human insulin.
 A:Reference number: A93144
 A:Accession: A93144
 A:Molecule type: protein
 A:Residues: 25-54; 90-110 <NIC>
 R:Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.
 J. Biol. Chem. 246, 1375-1386, 1971
 A:Title: Studies on human proinsulin. Isolation and amino acid sequence of the human par
 A:Reference number: A92075; MUID:71116410
 A:Accession: A92075
 A:Molecule type: protein
 A:Residues: 57-87 <OYE>
 R:Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
 Eur. J. Biochem. 20, 190-199, 1971
 A:Title: Amino acid sequence of the C-peptide of human proinsulin.
 A:Reference number: A91186; MUID:71257722
 A:Accession: A91186
 A:Molecule type: protein
 A:Residues: 57-87 <KOA>
 R:Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell
 Nature Genet. 4, 305-310, 1993
 A:Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment

A:Reference number: I58114; MUID:93364428
 A:Accession: I58114
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-59,63-110 <RES>
 A:Cross-references: GB:LI5440; NID:G307071; PIDN:AAA59179.1; PID:G307072
 R:Sieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.
 Helv. Chim. Acta 57, 2617-2621, 1974
 A:Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindung
 A:Reference number: A91636; MUID:75077277
 A:Contents: annotation; synthesis
 A:Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was i
 R:Naithani, V.K.
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
 A:Title: The synthesis of C-peptide of human proinsulin.
 A:Reference number: A91658; MUID:75040007
 A:Contents: annotation; synthesis of residues 57-87
 R:Geiger, R.; Jaeger, G.; Koenig, W.
 Chem. Ber. 106, 2347-2352, 1973
 A:Title: Synthesis of the complete sequence of human proinsulin C-peptide and its
 A:Reference number: A90914
 A:Contents: annotation; synthesis of residues 57-87
 R:Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.
 Biochem. J. 310, 869-874, 1995
 A:Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide j
 A:Reference number: S58661; MUID:96013185
 A:Contents: annotation; site-directed mutagenesis study of proteolytic processing
 C:Genetics:
 A:Gene: GDB:INS
 A:Cross-references: GDB:119349; OMIM:176730
 A:Map position: 11p15.5-11p15.5
 A:Introns: 63/1
 C:Superfamily: insulin
 C:Keywords: hormone; pancreas
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54/Domain: insulin chain B #status experimental <BCH>
 F:57-87/Domain: insulin #status experimental <MAT>
 F:90-110/Domain: connecting C peptide #status experimental <CPEP>
 F:91-96,43-109,95-100/Disulfide bonds: #status experimental
 F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 36.0%; Score 99; DB 1; Length 110;
 Best Local Similarity 69.0%; Pred. No. 2.3e-05;
 Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNFDPKT 51.
 II II III: I IIII IIIII: III
 DB 26 VNOHLCGSHLVEALYLVCGRGFFTPKT 54

RESULT 4

INSULIN precursor - green monkey
 C:Species: Cercopithecus aethiops (green monkey, grivet)
 C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
 C:Accession: B42179; A05232; S16494; S22056
 R:Seino, S.; Bell, G.I.; Li, W.H.
 Mol. Biol. Evol. 9, 193-203, 1992
 A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate
 A:Reference number: A42179; MUID:92219953
 A:Accession: B42179
 A:Molecule type: DNA
 A:Residues: 1-110 <SEI>
 A:Cross-references: EMBL:X61092; NID:G22808; PIDN:CAA43405.1; PID:G22809
 A:Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBI:95194)
 R:Peterson, J.D.; Nehrlich, S.; Oyer, P.E.; Steiner, D.F.
 J. Biol. Chem. 247, 4866-4871, 1972
 A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog pro
 A:Reference number: A92111; MUID:72258016
 A:Accession: A05232
 A:Molecule type: protein

A;Residues: 57-87 <PET>
C;Genetics:

A;Introns: 63/1
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-54/Domain: insulin chain B #status predicted <BCH>
F;25-54,90-110/Product: insulin #status predicted <MAT>
F;57-87/Domain: connecting peptide #status experimental <CPEP>
F;90-110/Domain: insulin chain A #status predicted <ACH>
F;31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 36.0%; Score 99; DB 2; Length 110;
Best Local Similarity 69.0%; Pred. No. 2.3e-05;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNPHLYGSHIREWLYLVAGRGFNFDPKT 51
|| || ||| : |||| ||||| : |||
Db 26 VNHLGSHLVEALYLVCGERGFFYTPKT 54

RESULT 5

A42179

insulin precursor - chimpanzee
C;Species: Pan troglodytes (chimpanzee)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C;Accession: A42179; S22058
R;Seino, S.; Bell, G.I.; Li, W.H.
Mol. Biol. Evol. 9, 193-203, 1992
A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of evolution in the primate lineage
A;Reference number: A42179; MUID:92219953
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-110 <SEI>
A;Cross-references: EMBL:X61089; NID:g38251; PIDN:CAA43403.1; PID:g38252
A;Note: sequence extracted from NCBI backbone (NCBIP:95067)
C;Genetics:
A;Introns: 63/1
C;Superfamily: insulin

Query Match 36.0%; Score 99; DB 2; Length 110;
Best Local Similarity 69.0%; Pred. No. 2.3e-05;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNPHLYGSHIREWLYLVAGRGFNFDPKT 51
|| || ||| : |||| ||||| : |||
Db 26 VNHLGSHLVEALYLVCGERGFFYTPKT 54

RESULT 6

JQ0178

insulin precursor - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 16-Jul-1999
C;Accession: JQ0178
R;Wetkam, W.; Gronenberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.
Gene 19, 179-183, 1982
A;Title: The nucleotide sequence of cDNA coding for preproinsulin from the primate Macaca fascicularis
A;Reference number: JQ0178; MUID:83080474
A;Accession: JQ0178
A;Molecule type: mRNA
A;Residues: 1-110 <MET>
A;Cross-references: GB:J000336; NID:g342121; PIDN:AAA36849.1; PID:g342122
C;Superfamily: insulin
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-54,90-110/Product: insulin #status predicted <MAT>
F;57-87/Domain: insulin chain B #status predicted <BCH>
F;55-89/Domain: insulin connecting C peptide #status predicted <CPT>
F;90-110/Domain: insulin chain A #status predicted <ACH>
F;31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 36.0%; Score 99; DB 2; Length 110;
Best Local Similarity 69.0%; Pred. No. 2.3e-05;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNPHLYGSHIREWLYLVAGRGFNFDPKT 51
|| || ||| : |||| ||||| : |||
Db 26 VNHLGSHLVEALYLVCGERGFFYTPKT 54

RESULT 7

INGS

insulin - goose
C;Species: Anser anser (domestic goose)
C;Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C;Accession: JC0007
R;Xu, Y.; Lin, N.; Zhang, Y.; Zhang, Y.
Kexue Tongbao 28, 966-968, 1983
A;Title: Isolation and sequence determination of goose insulin.
A;Reference number: JC0007
A;Accession: JC0007
A;Molecule type: protein
A;Residues: 1-30;31-51 <XUY>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCH>
F;31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 34.9%; Score 96; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 2.2e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 24 NFHLGSHIREWLYLVAGRGFNFDPKT 51
|| || ||| : |||| ||||| : |||
Db 3 NOHLGSHLVEALYLVCGERGFFYSPKT 30

RESULT 8

IPDK

insulin precursor - duck (tentative sequence)
C;Species: Anas platyrhynchos (domestic duck)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C;Accession: A01600; A91205
R;Markussen, J.; Sundby, F.
Int. J. Pept. Protein Res. 5, 37-48, 1973
A;Title: Duck insulin: Isolation, crystallization and amino acid sequence.
A;Reference number: A91766; MUID:74055140
A;Accession: A01600
A;Molecule type: protein
A;Residues: 1-30;61-81 <MAR>
A;Experimental source: Pekin breed
R;Markussen, J.; Sundby, F.
Eur. J. Biochem. 34, 401-408, 1973

A;Title: Isolation and amino-acid sequence of the C-peptide of duck proinsulin.
A;Reference number: A91205; MUID:73210102
A;Accession: A91205
A;Molecule type: protein
A;Residues: 33-58 <MAR2>
C;Comment: X's at positions 31-32 and 59-60 represent paired basic residues predicted

Query Match 34.9%; Score 96; DB 1; Length 81;

Best Local Similarity 67.9%; Pred. No. 3.8e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 24 NFHLGSHIREWLYLVAGRGFNFDPKT 51
DB 3 NOHLGSHLVREALYLVCGRGFFYSPK 30

RESULT 9

INHY

insulin - hamster
C:Species: Cricetinae gen. sp. (hamster)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A91456
R:Neelson, F.A.; Delcher, H.K.; Steinman, H.; Lebovitz, H.E.
Fed. Proc. 32, 300, 1973
A:Title: Structure of hamster insulin: comparison with a tumor insulin.
;Reference number: A91456

;Accession: A91456
A:Molecule type: protein
A:Residues: 1-30; 31-51 <NEE>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30,31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 34.5%; Score 95; DB 1; Length 51;
Best Local Similarity 65.5%; Pred. No. 3e-05;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNFDPKT 51
DB 2 VNOHLGSHLVREALYLVCGRGFFYTPKS 30

RESULT 10

INHY

insulin precursor - Selaphorus rufus (fragment)
C:Species: Selaphorus rufus
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 16-Jul-1999
C:Accession: I51221
R:Fau, L.; Gardner, P.; Chan, S.J.; Steiner, D.F.
Gen. Comp. Endocrinol. 91, 25-30, 1993
A:Title: Cloning and analysis of the gene encoding hummingbird proinsulin.
;Reference number: I51221; MUID:94010162

A:Accession: I51221
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-103 <FAN>
A:Cross-references: GB:S66612; NID:g451261; PIDN:RAC64211.1; PID:g451262
C:Genetics:
A:Introns: 59/1
C:Superfamily: insulin

Query Match 34.5%; Score 95; DB 2; Length 103;
Best Local Similarity 67.9%; Pred. No. 6.7e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNFDPK 50
DB 22 VNOHLGSHLVREALYLVCGRGFFYSPK 49

RESULT 11

INRB

insulin precursor - rabbit
N:Alternate names: preproinsulin
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 24-Apr-1984 #sequence_revision 23-Aug-1997 #text_change 18-Jun-1999

C:Accession: A53438; A01581
R;Devaskar, S.O.; Giddings, S.J.; Rajakumar, P.A.; Carnaghi, L.R.; Menon, R.K.; Zai
J. Biol. Chem. 269, 8445-8454, 1994
A:Title: Insulin gene expression and insulin synthesis in mammalian neuronal cells
A:Reference number: A53438; MUID:94179230
A:Accession: A53438
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-110 <DEV>
A:Cross-references: GB:U03610; NID:g467970; PIDN:AAA19033.1; PID:g467971
R:Smith, L.F.

Am. J. Med. 40, 662-666, 1966
A:Title: Species variation in the amino acid sequence of insulin.
A:Reference number: A90029; MUID:66160119
A:Accession: A01581

A:Molecule type: protein
A:Residues: 25-54; 90-110 <SMT>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54/Domain: insulin chain B #status experimental <BCH>
F:25-54,90-110/Product: insulin #status experimental <MAT>
F:57-87/Domain: connecting C peptide #status predicted <CPEP>
F:90-110/Domain: insulin chain A #status experimental <ACH>
F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 34.5%; Score 95; DB 1; Length 110;
Best Local Similarity 65.5%; Pred. No. 7.2e-05;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNFDPKT 51
DB 26 VNOHLGSHLVREALYLVCGRGFFYTPKS 54

RESULT 12

INHY

insulin precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999
C:Accession: I48166
R:Bell, G.I.; Sanchez-Pescador, R.
Diabetes 33, 297-300, 1984
A:Title: Sequence of a cDNA encoding Syrian hamster preproinsulin.
A:Reference number: I48166; MUID:84133036
A:Accession: I48166
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-110 <RES>
A:Cross-references: GB:M26328; NID:g191420; PIDN:AAA37089.1; PID:g305360
C:Superfamily: insulin

Query Match 34.5%; Score 95; DB 2; Length 110;
Best Local Similarity 65.5%; Pred. No. 7.2e-05;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNFDPKT 51
DB 26 VNOHLGSHLVREALYLVCGRGFFYTPKS 54

RESULT 13

INWHP

insulin - sperm whale
C:Species: Physeter catodon (sperm whale)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A93142; A90082
R:Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.
Nature 181, 1468-1469, 1958
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale p
A:Reference number: A93142

A:Accession: A93142
A:Molecule type: protein
A:Residues: 1-30;31-51 <ISH>
R: Harris, J. I.; Sanger, F.; Naughton, M. A.
Arch. Biochem. Biophys. 65, 427-428, 1956
A:Title: Species differences in insulin.
A:Reference number: A90082
A:Accession: A90082

A:Molecule type: protein
A:Residues: 1-30;31-51 <HAR>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30;31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 34.2%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 4e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNDPK 50
II II III: I IIII IIIII: II
Db 2 VNHLCGSHLVEALYLCGERGFFFTPK 29

RESULT 14

INWHF

insulin - finback whale (tentative sequence)
C:Species: Balaenoptera physalus (finback whale, common rorqual)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C:Accession: A91918
R:Hama, H.; Titani, K.; Sakaki, S.; Narita, K.
J. Biochem. 56, 285-293, 1964
A:Title: The amino acid sequence in fin-whale insulin.

A:Reference number: A91918
A:Accession: A91918

A:Molecule type: protein
A:Residues: 1-30;31-51 <HAM>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>
F:1-30;31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 34.2%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 4e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNDPK 50
II II III: I IIII IIIII: II
Db 2 VNHLCGSHLVEALYLCGERGFFFTPK 29

RESULT 15

INWHIS

insulin - sei whale
C:Species: Balaenoptera borealis (sei whale)
C:Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999
C:Accession: A01582
R:Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.
Nature 181, 1468-1469, 1958
A:Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin.
A:Reference number: A93142
A:Accession: A01582

A:Molecule type: protein
A:Residues: 1-30;31-51 <ISH>
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: insulin chain B #status experimental <BCH>

F:1-30,31-51/Product: insulin #status experimental <MAT>
F:31-51/Domain: insulin chain A #status experimental <ACH>
F:7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 34.2%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 4e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNDPK 50
II II III: I IIII IIIII: II
Db 2 VNHLCGSHLVEALYLCGERGFFFTPK 29

Search completed: August 23, 2002, 14:33:25
Job time: 128 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 14:32:27 ; Search time 10.36 Seconds
(without alignments)
190.608 Million cell updates/sec

Title: US-09-574-443-7
Perfect score: 275
Sequence: 1 NLVEQASTSQASLQIYNFD.....IREWLYVAGERGFNFDPKT 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	ID	Description
1	99	36.0	51	1	INS_ELEMA
2	99	36.0	110	1	INS_CERAE
3	99	36.0	110	1	INS_HUMAN
4	99	36.0	110	1	INS_MACFA
5	99	36.0	110	1	INS_PANTR
6	96	34.9	51	1	INS_ANGAN
7	96	34.9	81	1	INS_ANAPL
8	95	34.5	103	1	INS_SELRF
9	95	34.5	110	1	INS_CRILLO
10	95	34.5	110	1	INS_RABIT
11	94	34.2	51	1	INS_BALBO
12	94	34.2	51	1	INS_BALPH
13	94	34.2	51	1	INS_CAPHI
14	94	34.2	51	1	INS_DIDMA
15	94	34.2	51	1	INS_FELCA
16	94	34.2	86	1	INS_HORSE
17	94	34.2	105	1	INS_BOVIN
18	94	34.2	105	1	INS_SHEEP
19	94	34.2	108	1	INS_PIG
20	94	34.2	110	1	INS_CANFA
21	94	34.2	110	1	INS_PSAOB
22	93	33.8	108	1	INS_AOTTR
23	92	33.5	51	1	INS_ACOCA
24	91	33.1	51	1	INS_TRASC
25	91	33.1	107	1	INS_CHICK
26	90	32.7	50	1	INS_KATPE
27	90	32.7	51	1	INS_CAMDR
28	89	32.4	106	1	INS1_XENLA
29	88	32.0	106	1	INS2_XENLA
30	85	30.9	52	1	INS_ACIGU
31	84	30.5	52	1	INS_LEFSP
32	84	30.5	108	1	INS1_MOUSE
33	84	30.5	110	1	INS1_RAT

34	83	30.2	50	1	INS_MYOSC
35	83	30.2	50	1	INS_ONCSO
36	83	30.2	51	1	INS1_BATSP
37	83	30.2	51	1	INS2_THUTH
38	83	30.2	51	1	INS_CHIBR
39	83	30.2	51	1	INS_GADCA
40	83	30.2	51	1	INS_ORNAN
41	83	30.2	110	1	INS2_MOUSE
42	83	30.2	110	1	INS2_RAT
43	83	30.2	116	1	INS_LOPPI
44	82	29.8	51	1	INS_ANGRO
45	81	29.5	51	1	INS_HYSCR

ALIGNMENTS

RESULT 1
INS_ELEMA
ID INS_ELEMA STANDARD; PRT; 51 AA.
AC P01316;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Elephas maximus (Indian elephant).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.
OX NCBI_TaxID=9783;
RN [1]
RP MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.";
RL Am. J. Med. 40:662-666(1966).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THE SPECIES OF ELEPHANT IS NOT GIVEN, BUT IT IS
CC MOST PROBABLY THE INDIAN ELEPHANT (ELEPHAS MAXIMUS).
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR; A01584; INEL.
DR HSP; P01308; IBEN.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF000049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5752 MW; 9007B50CDB457D6D CRC64;

Query Match 36.0%; Score 99; DB 1; Length 51;
Best Local Similarity 69.0%; Pred. No. 2.8e-06;
Matches 20; Conservative 2; Mismatches 7; Indels 0;

OY 23 VNFHLYGSHIREWLYVAGERGFNFDPKT 51
|| || ||| : |||| |||| : |||
DB 2 VNOHLCGSHLVREALYLVCGERGFYFPKT 30

```

RESULT 2
INS_CERAE
ID INS_CERAE STANDARD; PRT; 110 AA.
AC P30407; P01309;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin precursor.
GN INS.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219953; PubMed=1560757;
RA Saino S., Bell G.I., Li W.;
T "Sequences of primate insulin genes support the hypothesis of a
t slower rate of molecular evolution in humans and apes than in
AT monkeys."
RL Mol. Biol. Evol. 9:193-203(1992).
RN [2]
RP SEQUENCE OF 57-87.
RX MEDLINE=72258016; PubMed=4626369;
RA Peterson J.D., Nehrllich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
RT dog proinsulin C-peptides by a semi-micro Edman degradation
RT procedure."
RL J. Biol. Chem. 247:4866-4871(1972).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X61092; CAA43405.1;
CC PIR; A05232; A05232.
CC PIR; S22056; S22056.
CC PIR; B42179; B42179.
CC HSSP; P01308; IBEN.
CC InterPro; IPR000739; Insulin_IGF_relaxin.
CC Pfam; PF00049; Insulin; 1.
CC PRINTS; PR00276; INSULIN.
CC PRINTS; PR00277; INSULIN.
CC SMART; SM00078; ILGF; 1.
CC PROSITE; PS00262; INSULIN; 1.
CC Insulin family; Hormone; Glucose metabolism; Signal.
CC SIGNAL 1 24
CC CHAIN 25 54 INSULIN B CHAIN.
CC PROPEP 57 87 C PEPTIDE.
CC CHAIN 90 110 INSULIN A CHAIN.
CC DISULFID 31 96 INTERCHAIN.
CC DISULFID 43 109 INTERCHAIN.
CC DISULFID 95 100
CC SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;

```

```

Query Match          36.0%; Score 99; DB 1; Length 110;
Best Local Similarity 69.0%; Pred. No. 6.8e-06;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

```

```

Oy 23 VNPHLYGSHIREWLYLVAGERGFFDPKT 51
   || || ||| | |||| ||||| : |||
Db 26 VNQHLCGSHLVEALYLVCGERGFFYTPKT 54

RESULT 3
INS_HUMAN
ID INS_HUMAN STANDARD; PRT; 110 AA.
AC P01308;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Insulin precursor.
GN INS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80120725; PubMed=6243748;
RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischer E.,
RA Goodman H.M.;
RT "Sequence of the human insulin gene."
RL Nature 284:26-32(1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=80236313; PubMed=6248962;
RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
RT "Genetic variation in the human insulin gene."
RL Science 209:612-615(1980).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80054779; PubMed=503234;
RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
RA Rutter W.J.;
RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin."
RL Nature 282:525-527(1979).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=80147417; PubMed=6927840;
RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
RT "Nucleotide sequence of human preproinsulin complementary DNA."
RL Science 208:57-59(1980).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=93364428; PubMed=8358440;
RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;
RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
RT kb segment of DNA spanning the insulin gene and associated VNTR."
RL Nat. Genet. 4:305-310(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE-Pancreas;
RA Strausberg R.;
RN Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1-59 FROM N.A.
RC TISSUE-Blood;
RA Fajardy I.I., Weill J.J., Stuckens C.C., Danze P.M.P.;
RT "Description of a novel RFLP diallelic polymorphism (-127 BsgI C/G)
RT within the 5' region of insulin gene."
RN Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 25-54 AND 90-110.
RA Nicol D.S.H.W., Smith L.F.;
RT "Amino-acid sequence of human insulin."
RL Nature 187:483-485(1960).
RN [9]
RP SEQUENCE OF 57-87.
RX MEDLINE=71116410; PubMed=5101771;
Oyer P.E., Cho S., Peterson J.D., Steiner D.F.;
RT "Studies on human proinsulin. Isolation and amino acid sequence of

```

RT the human pancreatic C-peptide.";
RL J. Biol. Chem. 246:1375-1386(1971).
RN [10]
RX SEQUENCE OF 57-87.
RA MEDLINE-71257722; PubMed-5560404;
RL Ko A., Smyth D.G., Markussen J., Sundby F.;
RT "The amino acid sequence of the C-peptide of human proinsulin.";
RN Eur. J. Biochem. 20:190-199(1971).
RX SYNTHESIS.
RA MEDLINE-75077277; PubMed-4443293;
RL Sieber P., Kamber B., Hartmann A., Joehl A., Rinkler B., Rittel W.;
RT "Total synthesis of human insulin under directed formation of the
RL disulfide bonds.";
RN Helv. Chim. Acta 57:2617-2621(1974).
RX SYNTHESIS OF 57-87.
RA MEDLINE-75040007; PubMed-4803504;
RL Naithani V.K.;
RT "Studies on polypeptides, IV. The synthesis of C-peptide of human
RL proinsulin.";
RN Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).
RX SYNTHESIS OF 65-69 AND 70-73.
RA MEDLINE-73161263; PubMed-4698555;
RL Geiger R., Volk A.;
RT "Synthesis of peptides with the properties of human proinsulin C
RL peptides (hc peptide). 3. Synthesis of the sequences 14-17 and 9-13
RL of human proinsulin C peptides.";
RN Chem. Ber. 106:199-205(1973).
RX SYNTHESIS OF 84-87.
RA MEDLINE-73161261; PubMed-4698553;
RL Geiger R., Jaeger G., Keonig W., Treuth G.;
RT "Synthesis of peptides with the properties of human proinsulin C
RL peptides (hc peptide). I. Scheme for the synthesis and preparation of
RL the sequence 28-31 of human proinsulin C peptide.";
RN Chem. Ber. 106:188-192(1973).
RX VARIANT LOS ANGELES.
RA MEDLINE-84016053; PubMed-6312455;
RL Haneda M., Chan S.J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;
RT "Studies on mutant human insulin genes: identification and sequence
RL analysis of a gene encoding [Ser24]insulin.";
RN Proc. Natl. Acad. Sci. U.S.A. 80:6366-6370(1983).
RX VARIANTS LOS ANGELES AND CHICAGO.
RA MEDLINE-84170233; PubMed-6424111;
RL Rubenstein A.H., Tager H.;
RT "Identification of a mutant human insulin predicted to contain a
RL serine-for-phenylalanine substitution.";
RN Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).
RX VARIANT PROVIDENCE.
RA MEDLINE-87175640; PubMed-3470784;
RL Chan S.J., Seino S., Gruppiso P.A., Schwartz R., Steiner D.F.;
RT "A mutation in the B chain coding region is associated with impaired
RL proinsulin conversion in a family with hyperproinsulinemia.";
RN Proc. Natl. Acad. Sci. U.S.A. 84:2194-2197(1987).
RX VARIANT WAKAYAMA.
RA MEDLINE-87058122; PubMed-3537011;
RL Sakura H., Iwamoto Y., Sakamoto Y., Kuzuya T., Hirata H.;
RT "Structurally abnormal insulin in a diabetic patient. Characterization
RL of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";
RN J. Clin. Invest. 78:1666-1672(1986).
RX VARIANT HIS-89.
RA MEDLINE-90317021; PubMed-2196279;
RL Barbeti F., Raben N., Kadowaki T., Cama A., Accili D., Gabbay K.H.,
RT Merenich J.A., Taylor S.I., Roth J.;
RN "Two unrelated patients with familial hyperproinsulinemia due to a

RT mutation substituting histidine for arginine at position 65 in the
RL proinsulin molecule: identification of the mutation by direct
RN sequencing of genomic deoxyribonucleic acid amplified by polymerase
RX chain reaction.";
RA J. Clin. Endocrinol. Metab. 71:164-169(1990).
RN [20]
RX VARIANT HIS-89.
RA MEDLINE-85261996; PubMed-4019786;
RL Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;
RT "Posttranslational cleavage of proinsulin is blocked by a point
RL mutation in familial hyperproinsulinemia.";
RN J. Clin. Invest. 76:378-380(1985).
RX VARIANT KYOTO.
RA MEDLINE-92291307; PubMed-1601997;
RL Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;
RT "A novel point mutation in the human insulin gene giving rise to
RL hyperproinsulinemia (proinsulin Kyoto).";
RN J. Clin. Invest. 89:1902-1907(1992).
RX STRUCTURE BY NMR.
RA MEDLINE-91104966; PubMed-2271664;
RL Hua Q.-X., Weiss M.A.;
RT "Toward the solution structure of human insulin: sequential 2D 1H NMR
RL assignment of a des-pentapeptide analogue and comparison with crystal
RL structure.";
RN Biochemistry 29:10545-10555(1990).
RX STRUCTURE BY NMR.
RA MEDLINE-91242467; PubMed-2036420;
RL Hua Q.-X., Weiss M.A.;
RT "Comparative 2D NMR studies of human insulin and des-pentapeptide
RL insulin: sequential resonance assignment and implications for protein
RL dynamics and receptor recognition.";
RN Biochemistry 30:5505-5515(1991).
RX STRUCTURE BY NMR.
RA MEDLINE-91265527; PubMed-1646635;
RL Hua Q.-X., Weiss M.A.;
RT "Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-
RL specific resonance assignments and effects of solvent composition.";
RN Biochim. Biophys. Acta 1078:101-110(1991).
RX STRUCTURE BY NMR.
RA MEDLINE-93059366; PubMed-1433291;
RL Joergensen A.M.M., Kristensen S.M., Led J.J., Balschmidt P.;
RT "Three-dimensional solution structure of an insulin dimer. A study of
RL the B9(Asp) mutant of human insulin using nuclear magnetic resonance,
RL distance geometry and restrained molecular dynamics.";
RN J. Mol. Biol. 227:1146-1163(1992).
RX STRUCTURE BY NMR OF VARIANT LOS-ANGELES.
RA MEDLINE-93133832; PubMed-8421693;
RL Hua Q.-X., Shoelson S.E., Inouye K., Weiss M.A.;
RT "Paradoxical structure and function in a mutant human insulin
RL associated with diabetes mellitus.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:582-586(1993).
RX STRUCTURE BY NMR.
RA MEDLINE-97383146; PubMed-9235985;
RL Chang X., Joergensen A.M., Bardrum P., Led J.J.;
RN Query Match 36.0%; Score 99; DB 1; Length 110;
Best Local Similarity 69.0%; Pred. No. 6.8e-06;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
Qy 23 VNFHLYGSHIREWLYLVAGERGFNFDPKT 51
||| ||| : |||| : |||
Db 26 VNHLCGSHLVEALYLVCGERGFFYTPKT 54
RESULT 4
INS_MACFA

INS_MACFA STANDARD; PRT; 110 AA.
 AC P30406; P01309;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Insulin precursor.
 GN INS.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83080474; PubMed=6184262;
 RA Wetekamp W., Groneberg J., Leineweber M., Wengenmayer F.,
 RA Winnacker E.-L.;
 RT "The nucleotide sequence of cDNA coding for preproinsulin from the
 RT primate Macaca fascicularis.";
 RT Gene 19:179-183(1982).
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; J00336; AAA36849.1; -
 DR PIR; J00178; J00178.
 DR HSP; P01308; IBEN.
 DR PIR; P000049; Insulin; 1.
 DR PRINTS; PR00276; INSULIN.
 DR PRINTS; PR00277; INSULIN.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 DR Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INTERCHAIN.
 FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100
 SQ SEQUENCE 110 AA; 11991 MW; 83C6E33A80A420F9 CRC64;
 Query Match 36.0%; Score 99; DB 1; Length 110;
 Best Local Similarity 69.0%; Pred. No. 6.8e-06;
 Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 OY 23 VNFHLYGSHIREWLYLVAGRGFNFDPKT 51
 DB 26 VNOHLCGSHLVYALVCGRGFFFTPKT 54
 RESULT 5
 INS_PANTR STANDARD; PRT; 110 AA.
 AC P30410;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Insulin precursor.
 GN INS.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92219953; PubMed=1560757;
 RA Seino S., Bell G.I., Li W.;
 RT "Sequences of primate insulin genes support the hypothesis of a
 RT slower rate of molecular evolution in humans and apes than in
 RT monkeys.";
 RT Mol. Biol. Evol. 9:193-203(1992).
 CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; X61089; CAA43403.1; -
 DR PIR; S22058; S22058.
 DR PIR; A42179; A42179.
 DR HSP; P01308; IBEN.
 DR InterPro; IPR000739; Insulin_IGF_relaxin.
 DR PIR; P00049; Insulin; 1.
 DR PRINTS; PR00276; INSULIN.
 DR PRINTS; PR00277; INSULIN.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 DR Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INTERCHAIN.
 FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100
 SQ SEQUENCE 110 AA; 12025 MW; 41EB8DF79837CEP5 CRC64;
 Query Match 36.0%; Score 99; DB 1; Length 110;
 Best Local Similarity 69.0%; Pred. No. 6.8e-06;
 Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 OY 23 VNFHLYGSHIREWLYLVAGRGFNFDPKT 51
 DB 26 VNOHLCGSHLVYALVCGRGFFFTPKT 54
 RESULT 6
 INS_ANSA STANDARD; PRT; 51 AA.
 AC P07454; Q10995;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Insulin.
 GN INS.
 OS Anser anser anser (Western graylag goose), and
 OS Cairina moschata (Muscovy duck).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

RX MEDLINE=73210102; PubMed=4715652;
RA Markussen J., Sundby F.;
RT "Isolation and amino-acid sequence of the C-peptide of duck
RL proinsulin.";
RL Eur. J. Biochem. 34:401-408(1973).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -1- CAUTION: X'S AT POSITIONS 31-32 AND 59-60 REPRESENT PAIRED BASIC
CC RESIDUES ASSUMED BY HOMOLOGY TO BE PRESENT IN THE PRECURSOR
CC MOLECULE.
CC PIR; A01600; IPDK.
DR HSP; P01308; IHIS.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT PROPEP 33 58 C PEPTIDE.
FT CHAIN 61 81 INSULIN A CHAIN.
FT DISULFID 7 67 INTERCHAIN.
FT DISULFID 19 80 INTERCHAIN.
FT DISULFID 66 71
FT SEQUENCE 81 AA; 9100 MW; 6EA8A271F099DA91 CRC64;
SQ

Query Match 34.98; Score 96; DB 1; Length 81;
Best Local Similarity 67.98; Pred. No. 1.2e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps

QY 24 NFHLGSHIREWLYLVAGERGFFNFDPKT 51
DB 3 NQHLGSHLVEALYLVCGERGFFYSPTK 30
1 11 111: 1 1111 11111: 111
3 NQHLGSHLVEALYLVCGERGFFYSPTK 30

RESULT 8
INS_SELFR ID INS_SELFR STANDARD; PRT; 103 AA.
AC P51463;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Insulin precursor (Fragment).
GN INS.
OS Selasphorus rufus (Hummingbird).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauaria; Aves; Neognathae; Trochiliformes; Trochilidae;
OC Selasphorus.
ON NCBI_TaxID=29060;
OX [1]
RN SEQUENCE FROM N.A.
RA MEDLINE=94010162; PubMed=8405887;
RX Fan L., Gardner P., Chan S.J., Steiner D.F.;
RT "Cloning and analysis of the gene encoding hummingbird proinsulin.";
RL Gen. Comp. Endocrinol. 91:25-30(1993).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collabora

```

 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U03610; AAA19033.1; .
 DR EMBL; M61153; AAA17540.1; .
 DR PIR; A01581; INRB.
 DR HSP; P01308; IXDA.
 DR InterPro; IPR000739; Insulin_IGF_relaxin.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00276; INSULINA.
 DR PRINTS; PR00277; INSULINE.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 DR Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INTERCHAIN.
 FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100
 FT CONFLICT 83 83 E -> Y (IN REF. 3).
 SQ SEQUENCE 110 AA; 11838 MW; 82D2975B85D77FA8 CRC64;

Query Match 34.5%; Score 95; DB 1; Length 110;
 Best Local Similarity 65.5%; Pred. No. 2.2e-05;
 Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYLVAGRGFNFDPKT 51
 II III IIII I IIII IIII I I I
 DB 26 VNHLCGSHLVEALYLVCGRGFFFTPKS 54

RESULT 11
 INS_BALBO
 ID INS_BALBO STANDARD; PRT; 51 AA.
 AC P01314;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Insulin.
 GN INS.
 OS Balaenoptera borealis (Sei whale).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balaenopteridae; Balaenoptera.
 OX NCBI_TaxID=9768;
 RN [1]
 RP SEQUENCE.
 RA Ishihara Y., Saito T., Ito Y., Fujino M.;
 RT "Structure of sperm- and sei-whale insulins and their breakdown by
 RT whale pepsin.";
 RL Nature 181:1468-1469(1958).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR PIR; A01582; INWH1.
 DR HSP; P01315; 9INS.
 DR InterPro; IPR000739; Insulin_IGF_relaxin.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00276; INSULINA.
 DR SMART; PR00276; INSULIN; 1.
 DR PRINTS; PR00277; INSULINE.

DR PRINTS; PR00277; INSULINE.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;
 Query Match 34.2%; Score 94; DB 1; Length 51;
 Best Local Similarity 67.9%; Pred. No. 1.2e-05;
 Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 23 VNFHLYGSHIREWLYLVAGRGFNFDPK 50
 II III IIII I IIII IIII I I I
 DB 2 VNHLCGSHLVEALYLVCGRGFFFTPK 29
 RESULT 12
 INS_BALPH
 ID INS_BALPH STANDARD; PRT; 51 AA.
 AC P01312;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Insulin.
 GN INS.
 OS Balaenoptera physalus (Finback whale) (Common rorqual), and
 OS Physeter catodon (Sperm whale) (Physeter macrocephalus).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balaenopteridae; Balaenoptera.
 OX NCBI_TaxID=9770, 9755;
 RN [1]
 RP PARTIAL SEQUENCE.
 RC SPECIES-B.physalus;
 RA Hama H., Titani K., Sakaki S., Narita K.;
 RT "The amino acid sequence in fin-whale insulin.";
 RL J. Biochem. 56:285-293(1964).
 RN [2]
 RP SEQUENCE.
 RC SPECIES-P.catodon;
 RA Ishihara Y., Saito T., Ito Y., Fujino M.;
 RT "Structure of sperm- and sei-whale insulins and their breakdown by
 RT whale pepsin.";
 RL Nature 181:1468-1469(1958).
 RN [3]
 RP SEQUENCE.
 RC SPECIES-P.catodon;
 RA Harris J.I., Sanger F., Naughton M.A.;
 RT "Species differences in insulin.";
 RL Arch. Biochem. Biophys. 65:427-438(1956).
 CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
 CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
 CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
 CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
 CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR PIR; A91918; INWHF.
 DR PIR; A93142; INWHF.
 DR HSP; P01315; 9INS.
 DR InterPro; IPR000739; Insulin_IGF_relaxin.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00276; INSULINA.
 DR PRINTS; PR00277; INSULINE.
 DR SMART; SM00078; ILGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT CHAIN 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5766 MW; 9007B514691A/CDD CRC64;

Query Match 34.2%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 1.2e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 23 VNFHLYGSHIREWLYLVAGRGFNFDPK 50
II III: I IIII IIII: II
DB 2 VNOHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 13
INS_CAPHI STANDARD; PRT; 51 AA.
AC P01319;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE.

RA MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.";
RL Am. J. Med. 40:662-666(1966).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC PIR; A01586; INGT.
CC HSP; P01315; 9INS.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5692 MW; 9007B50CDB4E7DDD CRC64;

Query Match 34.2%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 1.2e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 23 VNFHLYGSHIREWLYLVAGRGFNFDPK 50
II III: I IIII IIII: II
DB 2 VNOHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 14

INS_DIDMA STANDARD; PRT; 51 AA.
AC P18109;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.

OS Didelphis marsupialis virginiana (North American opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.
OX NCBI_TaxID=9267;
RN [1]
RP SEQUENCE.

RC TISSUE=Pancreas;
RX MEDLINE=90160042; PubMed=2695899;
RA Yu J.-H., Eng J., Rattan S., Yalow R.S.;
RT "Opossum insulin, glucagon and pancreatic polypeptide: amino acid
sequences.";
RL Peptides 10:1195-1197(1989).

CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.

CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC PIR; JQ0362; JQ0362.
DR PIR; JQ0363; JQ0363.

DR HSP; P01317; 2INS.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5732 MW; 9007B8BAE4BDEDD CRC64;

Query Match 34.2%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 1.2e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 23 VNFHLYGSHIREWLYLVAGRGFNFDPK 50
II III: I IIII IIII: II
DB 2 VNOHLCGSHLVEALYLVCGRGFFFTPK 29

RESULT 15

INS_FELCA STANDARD; PRT; 51 AA.
AC P06306;
DT 01-JAN-1988 (Rel. 06, Created)

DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin.
GN INS.

OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE.

RX MEDLINE=86214076; PubMed=3518635;
RA Hallden G., Gafvelin G., Mutt V., Joernvall H.;
RT "Characterization of cat insulin."
RL Arch. Biochem. Biophys. 247:20-27(1986).
CC -I- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -I- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR; A01588; INCT.
DR HSP; P01315; 9INS.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
-KW Insulin family; Hormone; Glucose metabolism.
CHAIN 1 30 INSULIN B CHAIN.
NON_CONS 30 31
CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5745 MW; 9007B5096A0A7DDD CRC64;

Query Match 34.2%; Score 94; DB 1; Length 51;
Best Local Similarity 67.9%; Pred. No. 1.2e-05;
Matches 19; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 23 VNFHLYGSHIREWLYVAGERGFNEDPK 50
|| || ||| : |||| |||| : ||
Db 2 VNHLCGSHLYEALYVCGERGFYTPK 29

Search completed: August 23, 2002, 14:36:07
Job time: 220 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 14:32:07 ; Search time 25.36 Seconds
(without alignments)
347.900 Million cell updates/sec

Title: US-09-574-443-7
Perfect score: 275
Sequence: 1 NLVEQASTSQASLYQIYNFD.....IREWLVLVAGERGFNFDPKT 51

Scoring table: BLOSUM62
Gapop' 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	95	34.5	110	11	Q91X13 spermophilu
2	92	33.5	111	13	Q98TB0 chitalla chi
3	87	31.6	87	13	Q98TA9 gnathonemus
4	85	30.9	110	13	Q98TA8 pantodon bu
5	85	30.9	110	13	Q90ZY1 hiodon alos
6	83	30.2	91	13	Q98TB2 ambloplites
7	81.5	29.6	106	13	Q91BQ7 rana pipien
8	74	26.9	108	13	Q98TB1 catostomus
9	74	26.9	108	13	Q98DE5 brachydanio
10	74	26.9	108	13	Q90ZN4 catla catla
11	73	26.5	111	13	Q98TA7 osteoglossu
12	68	24.7	299	16	Q9KBM0 bacillus ha
13	63.5	23.1	539	12	Q68537 human calic
14	63	22.9	1462	16	O34313 bacillus su
15	62.5	22.7	539	12	Q66287 calicivirid
16	60.5	22.0	539	12	Q917Y6 norwalk-lik

17	59	21.5	151	8	Q9TCV1	Q9TCV1 metagonimus
18	59	21.5	374	2	Q9F8S3	Q9F8S3 thermus fil
19	59	21.5	664	4	Q9UIL7	Q9UIL7 homo sapien
20	59	21.5	1530	11	Q9WUG5	Q9WUG5 rattus norv
21	59	21.5	1531	11	O88279	O88279 rattus norv
22	59	21.5	1531	11	Q9WVB5	Q9WVB5 mus musculu
23	59	21.5	1534	4	O75093	O75093 homo sapien
24	58.5	21.3	189	16	Q992W5	Q992W5 streptococc
25	58.5	21.3	358	8	Q32253	Q32253 gilia austr
26	58.5	21.3	358	8	Q32268	Q32268 gilia carui
27	58.5	21.3	358	8	Q32356	Q32356 gilia splen
28	58.5	21.3	539	12	Q91323	Q91323 human calic
29	58	21.1	546	10	O82217	O82217 arabidopsis
30	57.5	20.9	95	16	Q928D3	Q928D3 listeria in
31	57.5	20.9	360	8	Q32809	Q32809 polemonium
32	57.5	20.9	360	8	Q9THX7	Q9THX7 polemonium
33	57	20.7	148	8	Q9TCV2	Q9TCV2 metagonimus
34	57	20.7	303	5	O9VB24	O9VB24 drosophila
35	57	20.7	397	10	Q92F65	Q92F65 lycopersico
36	57	20.7	609	16	Q9KNJ4	Q9KNJ4 vibrio chol
37	56.5	20.5	290	2	Q52614	Q52614 proteus vul
38	56.5	20.5	358	8	Q32181	Q32181 eriastrum d
39	56.5	20.5	358	8	Q32287	Q32287 gilia filif
40	56.5	20.5	358	8	Q32290	Q32290 gilia hutch
41	56.5	20.5	358	8	Q32442	Q32442 ipomopsis a
42	56.5	20.5	358	8	Q32464	Q32464 ipomopsis p
43	56.5	20.5	358	8	Q32300	Q32300 gilia leptu
44	56.5	20.5	358	8	Q32538	Q32538 langloisia
45	56.5	20.5	358	8	Q32826	Q32826 phlox graci

ALIGNMENTS

RESULT 1

Q91X13 ID Q91X13 PRELIMINARY; PRT; 110 AA.
AC Q91X13;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE INSULIN.
OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Scluridae; Sciurinae;
OC Spermophilus.
OX NCBI_TaxID=43179;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Tredrea M.M., Buck M.J., Guhaniyogi J., Squire T.L., Andrews M.T.;
RT "Regulation of PDK4 expression in a hibernating mammal.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY038604; AAK72558.1; -
SQ SEQUENCE 110 AA; 12004 MW; 4511768D6622BEE5 CRC64;

Query Match 34.5%; Score 95; DB 11; Length 110;
Best Local Similarity 65.5%; Pred. No. 6.1e-05;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLVLVAGERGFNFDPKT 51
||| ||| : ||| ||| : |||

DB 26 VNHLCGSHLVLEALYVCGERGFYTPKS 54

RESULT 2

Q98TB0 ID Q98TB0 PRELIMINARY; PRT; 111 AA.
AC Q98TB0;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE PREPROINSULIN (FRAGMENT).
OS Chitalla chitalla.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Notopteridae; Chitalla.
OX NCBI_TaxID=112163;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21203577; PubMed=11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF199586; AAK28710.1; -.
DR HSSP; P01308; ILPH.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12483 MW; 247CA4431376329F CRC64;

Query Match 33.5%; Score 92; DB 13; Length 111;
Best Local Similarity 66.7%; Pred. No. 0.00015;
Matches 18; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 24 NPHLYGSHIREWLYLVAGRGFNFDPK 50
| |||||: ||||| ||||| ||||| : |||||
Db 26 NQHLGSHLVEALFLVCGERGFFNPK 52

RESULT 3
Q98TA9 PRELIMINARY; PRT; 87 AA.
AC Q98TA9; (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PREPROINSULIN (FRAGMENT).
OS Gnathonemus petersi.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Mormyridae; Gnathonemus.
OX NCBI_TaxID=42645;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21203577; PubMed=11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF199587; AAK28711.1; -.
DR HSSP; P01308; IHS.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
FT NON_TER 87 87
SQ SEQUENCE 87 AA; 9874 MW; FF448ED35D2453F5 CRC64;

Query Match 31.6%; Score 87; DB 13; Length 87;
Best Local Similarity 65.4%; Pred. No. 0.00054;
Matches 17; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 26 HLHGSHIREWLYLVAGRGFNFDPK 51

Db 5 HLCGSHLVEALFLVCGERGFFNPD 30
|| ||||: ||||| ||||| ||||| : |||||
RESULT 4
Q98TA8 PRELIMINARY; PRT; 110 AA.
AC Q98TA8; (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PREPROINSULIN.
OS Pantodon buchholzi (Butterflyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Pantodontidae; Pantodon.
OX NCBI_TaxID=8276;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21203577; PubMed=11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-58(2001).
CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR EMBL; AF199588; AAK28712.1; -.
DR HSSP; P01308; IHS.
DR InterPro; IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
SQ SEQUENCE 110 AA; 12324 MW; BDECCD659D872E06 CRC64;

Query Match 30.9%; Score 85; DB 13; Length 110;
Best Local Similarity 57.7%; Pred. No. 0.0013;
Matches 15; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 26 HLHGSHIREWLYLVAGRGFNFDPK 51
|| ||||: ||||| ||||| ||||| : |||||
Db 28 HLCGSHLVDALYVCGEKGFYQPK 53

RESULT 5
Q90ZY1 PRELIMINARY; PRT; 110 AA.
AC Q90ZY1; (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PREPROINSULIN (FRAGMENT).
OS Hiodon alosoides (goldeye).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
OC Osteoglossiformes; Hiodontidae; Hiodon.
OX NCBI_TaxID=54904;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21203577; PubMed=11306171;
RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RT "Molecular cloning of preproinsulin cDNAs from several
osteoglossomorphs and a cyprinid."
RL Mol. Cell. Endocrinol. 174:51-58(2001).
DR EMBL; AF282408; AAK54684.1; -.
FT NON_TER 110 110
SQ SEQUENCE 110 AA; 12343 MW; BDECCD7703E52E06 CRC64;

Query Match 30.9%; Score 85; DB 13; Length 110;
Best Local Similarity 57.7%; Pred. No. 0.0013;
Matches 15; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OC	Osteoglossiformes; Osteoglossidae; Osteoglossum.
NCBI	NCBI_TaxID=109271;
OX	[1]
RN	SEQUENCE FROM N.A.
RP	MEDLINE=21203577; PubMed=11306171;
RX	Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
RY	"Molecular cloning of preproinsulin cDNAs from several
RT	osteoglossomorphs and a cyprinid.";
RT	Mol. Cell. Endocrinol. 174:51-58(2001).
CC	- I- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC	- I- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC	EMBL: AF195889; AAK28713.1; -
DR	HSSP; P01315; IMPJ.
DR	InterPro: IPR000739; Insulin_IGF_relaxin.
DR	Pfam: PF000049; Insulin. 1.
DR	PRINTS; PR00277; INSULINB.
DR	SMART; SM00078; IIGF; 1.
DR	PROSITE; PS00262; INSULIN; 1.
DR	NON_TER 111 111
DR	SEQUENCE 111 AA 12491 MW: AC9E19P2D4866D20 CRC64:
SO	

Query Match 26.5%; Score 73; DB 13; Length 111;
Best Local Similarity 52.0%; Pred. No. 0.05;
Matches 13; Conservative 6; Mismatches 6; Indels

Qy	27	LYGSHIREWLVLVAGERGFNFDPKT	51
		: : : :	
Db	30	LCGSHLVDALYMVCGDRGFFFYSPKS	54

Q90ZN4
ID Q90ZN4
PRELIMINARY;
PRT; 108 AA.

Q9KBM0	PRELIMINARY;	PRT;	299 AA.
AC	Q9KBM0		
AD	Q9KBM0;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DDT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DDT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	TRANSCRIPTIONAL REGULATOR (ARAC/XLIS FAMILY).		
DE	BH1906.		
GN	Bacillus halodurans.		
OS	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OS	Bacillus/Staphylococcus group; Bacillus.		
OX	NCBI_TaxID=86665;		
OX	{1}		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=C-125 / JCM 9153;		
RP	MEDLINE=20512582; PubMed=11058132;		
RR	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki		
RR	Fuji F., Hirajima C., Nakamura Y., Ogasawara N., Kuha		
RR	Horikoshi K.;		
RR	"Complete genome sequence of the alkaliphilic bacter		
RFT	halodurans and genomic sequence comparison with Bac		
RFT	Nucleic Acids Res. 28:4317-4331(2000).		
RL	-!- SIMILARITY: BELONGS TO THE ARAC/XLIS FAMILY OF		
RL	REGULATORS.		
CC			

Query Match 26.9%; Score 74; DB 13; Length 108;
Best Local Similarity 56.0%; Pred. No. 0.036;
Matches 14; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

26 HLYGSHIREWLYLVAGERGFNFDPK 50
 || ||| : |||| | || : ||
 28 HLCGSHLVDAALVYCGPTGFEYNPK 52

Q98TA7
ID Q98TA7
PRELIMINARY: PRT: 111 AA.

Query Match 24.7%; Score 68; DB 16; Length 299;
Best Local Similarity 30.8%; Pred. No. 0.7;
Matches 12; Conservative 8; Mismatches 19; Indels

QY 12 SLYYINFDNDVNFHLYGSHIREWLYLVAGERGFNDFPK 50
 :::|:| ||:| ||| |

```

Db 20 NVYRKVEGVQVHCHWHHEWYMEAGEAEFQIGTK 58
RESULT 13
Q68537 PRELIMINARY; PRT; 539 AA.
ID Q68537
AC Q68537;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE CAPSID.
OS human calicivirus isolates, and
OS Camberwell virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Norwalk-like viruses.
OX NCBI_TaxID=150080, 94433;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CAMBERWELL;
RX MEDLINE=96310230; PubMed=8732863;
RA Cauchi M.R., Doultree J.C., Marshall J.A., Wright P.J.;
RA "Molecular characterization of Camberwell virus and sequence variation
in ORP3 of small round-structured (Norwalk-like) viruses.";
J. Med. Virol. 49:70-76(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HU/NLV/CAMBERWELL/101922/94/AUS;
RA Seah E.L., Marshall J.A., Wright P.J.;
RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; U46500; AAC54893.1; -
DR EMBL; AF145896; AAC33961.1; -
DR InterPro; IPR004005; Calici_coat.
DR Pfam; PF00915; Calici_coat; 1.
SQ SEQUENCE 539 AA; 58869 MW; B4AFAB9C398A28E5 CRC64;

Query Match 23.1%; Score 63.5; DB 12; Length 539;
Best Local Similarity 38.1%; Pred. No. 5.3;
Matches 16; Conservative 7; Mismatches 16; Indels 3; Gaps 2;

QY 8 TSQASLYQIYFNFDNDVNFHLYGSHIREWLYLVAGERGFNFD 49
|:|:| | | | | | | | | | | | | | | | | | | | |
Db 275 TTQLSAVNICNFRGVDVTHVGS--DYTMNLASQNSWYDP 313

RESULT 14
Q34313 PRELIMINARY; PRT; 1462 AA.
ID Q34313
AC Q34313;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE YFKN PROTEIN.
GN YFKN.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Gulseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

```

```

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Rapoport G., Rey M., Reynolds S.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rose M., Sadale Y.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RA "The complete genome sequence of the gram-positive bacterium Bacillus
subtilis.";
RT subtitlis.;
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AC327;
RA Sekiguchi J., Yamamoto H., Uchiyama S., Fajar A.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RL EMBL; Z99108; CAB12613.1; -
DR EMBL; D83967; BAA23404.1; -
DR HSSP; P07024; 2USH.
DR InterPro; IPR002224; 5_nucleotidase.
DR InterPro; IPR000934; Ser_thr_phosphatse.
DR Pfam; PF01009; 5_nucleotidase; 2.
DR Pfam; PF02872; 5_nucleotidase; 2.
DR PROSITE; PS00785; 5_NUCLEOTIDASE_1; 2.
DR PROSITE; PS00786; 5_NUCLEOTIDASE_2; 1.
KW Complete proteome.
SQ SEQUENCE 1462 AA; 159705 MW; EE769201F34DA84C CRC64;

Query Match 22.9%; Score 63; DB 16; Length 1462;
Best Local Similarity 28.3%; Pred. No. 19;
Matches 15; Conservative 8; Mismatches 10; Indels 20; Gaps 2;

QY 16 IYFNDVNF-HLYGSHIREWLYLVAGE-----RGFNFD 48
|:|:| | | | | | | | | | | | | | | | | | | | |
Db 455 LYLDNTVOIVKLTGSEVKDLEMSAGQFNQIDPAKGGDQALLNFRSINFD 507

RESULT 15
Q66287 PRELIMINARY; PRT; 539 AA.
ID Q66287
AC Q66287;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE CAPSID PROTEIN.
OS Caliciviridae.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=11974;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL VIRUS;
RX MEDLINE=94321977; PubMed=8046390;
RA Green S.M., Dingle K.E., Lambden P.R., Caul E.O., Ashley C.R.,
RA Clarke I.N.;
RA "Human enteric Caliciviridae: a new prevalent small round-structured
group defined by RNA-dependent RNA polymerase and capsid diversity.";
J. Gen. Virol. 75:1883-1888(1994).
RL EMBL; X76716; CAA54134.1; -
DR EMBL; X76716; CAA54134.1; -

```

DR InterPro: IPR004005; Calici_coat.
 DR Pfam: PF00915; Calici_coat; 1.
 SQ SEQUENCE 539 AA; 58801 MW; 5093AB44F2C5E582 CRC64;

Query Match 22.7%; Score 62.5; DB 12; Length 539;
 Best Local Similarity 38.1%; Pred. No. 7.2; 16; Indels 3; Gaps 2;
 Matches 16; Conservative 7; Mismatches 16; Indels 3; Gaps 2;

OY 8 TSQASLYQIYNFDNDFHLYGSHIREWLYIVAGERGFD 49

Db 275 TTQLSAVNICFRGDYT-HIAGSH--DYTMNLASQNSNYDP 313

Search completed: August 23, 2002, 14:35:51
 Job time: 224 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 23, 2002, 14:30:07 ; Search time 13.01 Seconds
(without alignments)
95.750 Million cell updates/sec

Title: US-09-574-443-7

Perfect score: 275

Sequence: 1 NLVQASTSQSLYQIYNFD.....IREWLYVAGERGFNFDPKT 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	54.2	51	1	US-07-707-542E-2
2	144	52.4	51	1	US-07-707-542E-1
3	140	50.9	51	1	US-07-707-542E-3
4	128	46.5	51	1	US-07-707-542E-5
5	126	45.8	51	1	US-07-707-542E-4
6	122.5	44.5	86	1	US-07-715-183C-1
7	101	36.7	96	1	US-08-160-376A-5
8	101	36.7	96	1	US-08-389-487-8
9	99	36.0	30	1	US-08-212-696-2
10	99	36.0	30	1	US-08-233-617-2
11	99	36.0	30	1	US-08-160-376A-2
12	99	36.0	30	1	US-08-301-838-2
13	99	36.0	30	1	US-08-389-487-5
14	99	36.0	30	2	US-08-508-664-7
15	99	36.0	30	2	US-08-353-476-86
16	99	36.0	30	2	US-08-353-476-88
17	99	36.0	30	2	US-08-484-219-6
18	99	36.0	30	2	US-08-992-676-2
19	99	36.0	30	2	US-09-134-836-2
20	99	36.0	30	3	US-08-750-391-4
21	99	36.0	30	4	US-09-255-668-1
22	99	36.0	30	4	US-09-099-307-2
23	99	36.0	30	4	US-08-900-574-2
24	99	36.0	30	5	PCT-US94-04179-3
25	99	36.0	30	6	5169865-5
26	99	36.0	31	1	US-08-389-487-10
27	99	36.0	31	2	US-08-992-676-5

28	99	36.0	32	1	US-08-389-487-9	Sequence 9, Appli
29	99	36.0	32	2	US-08-992-676-4	Sequence 4, Appli
30	99	36.0	32	2	US-09-134-836-6	Sequence 6, Appli
31	99	36.0	36	1	US-08-030-731A-34	Sequence 34, Appli
32	99	36.0	53	1	US-08-233-617-3	Sequence 3, Appli
33	99	36.0	53	1	US-08-233-617-4	Sequence 4, Appli
34	99	36.0	53	4	US-08-900-574-4	Sequence 4, Appli
35	99	36.0	53	4	US-08-981-988A-42	Sequence 42, Appli
36	99	36.0	54	4	US-08-900-574-9	Sequence 9, Appli
37	99	36.0	55	4	US-08-900-574-6	Sequence 6, Appli
38	99	36.0	55	4	US-08-900-574-8	Sequence 8, Appli
39	99	36.0	56	1	US-08-160-376A-7	Sequence 7, Appli
40	99	36.0	56	1	US-08-389-487-11	Sequence 11, Appli
41	99	36.0	57	1	US-08-030-731A-44	Sequence 44, Appli
42	99	36.0	63	1	US-08-160-376A-6	Sequence 6, Appli
43	99	36.0	65	4	US-08-900-574-3	Sequence 3, Appli
44	99	36.0	66	1	US-08-291-060B-5	Sequence 5, Appli
45	99	36.0	66	4	US-08-900-574-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-07-707-542E-2
; Sequence 2, Application US/07707542E
; Patent No. 5242900
; GENERAL INFORMATION:
; APPLICANT: Albisser, A. Michael
; TITLE OF INVENTION: Treatment of Diabetes using
; TITLE OF INVENTION: Phosphorylated Insulin
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Queen's University
; STREET: PARTEQ Innovations
; CITY: Kingston
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 720K floppy
; COMPUTER: PC-AT compatible
; OPERATING SYSTEM: MSDOS 5.0
; SOFTWARE: Word for Windows 1.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07707,542E
; FILING DATE: 19910530
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hicks, Richard J.
; REGISTRATION NUMBER: 24665
; REFERENCE/DOCKET NUMBER: Q1341
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 545-2067
; TELEFAX: (613) 545-8853
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 residues
; TYPE: AMINO ACIDS
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: no
; ORIGINAL SOURCE:
; ORGANISM: Human pancreas
; FEATURE:
; OTHER INFORMATION: Peptide has two subunits. Second
; OTHER INFORMATION: subunit starts at position 22. Cys-Cys cross links
; OTHER INFORMATION: between positions 6 and 11, 7 and 28, and 20 and 40.
; OTHER INFORMATION: One or more of the Serine and/or Threonine residues
; OTHER INFORMATION: are phosphorylated as O-phosphates.
US-07-707-542E-2

Query Match 54.2% Score 149; DB 1; Length 51;

Best Local Similarity 62.0%; Pred. No. 4.2e-14;
Matches 31; Conservative 5; Mismatches 14; Indels 0; Gaps 0;
Qy 2 LVEQASTQASLYQIYNFNDVNFHLYGSHIREWLYLVAGERGFNDFPK 51
:111 11 1111: 1: 11 11 111: 1 111 11111: 111
Db 2 IVEQCTSTCSLYQLENYCNFVNHQHLGSHLVREALYLVCGRGFFFTPK 51

RESULT 2
US-07-707-542E-1
; Sequence 1, Application US/07707542E
; Patent No. 5242900
; GENERAL INFORMATION:
; APPLICANT: Albisser, A. Michael
; TITLE OF INVENTION: Treatment of Diabetes using
; TITLE OF INVENTION: Phosphorylated Insulin
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Queen's University
; STREET: PARTEQ Innovations
; CITY: Kingston
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 720K floppy
; COMPUTER: PC-AT compatible
; OPERATING SYSTEM: MSDOS 5.0
; SOFTWARE: Word for Windows 1.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07707,542E
; FILING DATE: 19910530
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hicks, Richard J.
; REGISTRATION NUMBER: 24665
; REFERENCE/DOCKET NUMBER: Q1341
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 545-2067
; TELEFAX: (613) 545-6853
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 residues
; TYPE: AMINO ACIDS
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: no
; ORIGINAL SOURCE:
; ORGANISM: Pig pancreas
; FEATURE:
; OTHER INFORMATION: Peptide has two subunits. Second
; OTHER INFORMATION: subunit starts at position 22. Cys-Cys cross links
; OTHER INFORMATION: between positions 6 and 11, 7 and 28, and 20 and 40.
; OTHER INFORMATION: One or more of the Serine and/or Threonine residues
; OTHER INFORMATION: are phosphorylated as O-phosphates.
US-07-707-542E-1

Query Match 52.4%; Score 144; DB 1; Length 51;
Best Local Similarity 61.2%; Pred. No. 2.1e-13;
Matches 30; Conservative 5; Mismatches 14; Indels 0; Gaps 0;
Qy 2 LVEQASTQASLYQIYNFNDVNFHLYGSHIREWLYLVAGERGFNDFPK 50
:111 11 1111: 1: 11 11 111: 1 111 11111: 111
Db 2 IVEQCTSTCSLYQLENYCNFVNHQHLGSHLVREALYLVCGRGFFFTPK 50

RESULT 3
US-07-707-542E-3
; Sequence 3, Application US/07707542E
; Patent No. 5242900
; GENERAL INFORMATION:
; APPLICANT: Albisser, A. Michael
; TITLE OF INVENTION: Treatment of Diabetes using

; TITLE OF INVENTION: Phosphorylated Insulin
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Queen's University
; STREET: PARTEQ Innovations
; CITY: Kingston
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 720K floppy
; COMPUTER: PC-AT compatible
; OPERATING SYSTEM: MSDOS 5.0
; SOFTWARE: Word for Windows 1.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07707,542E
; FILING DATE: 19910530
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hicks, Richard J.
; REGISTRATION NUMBER: 24665
; REFERENCE/DOCKET NUMBER: Q1341
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 545-2067
; TELEFAX: (613) 545-6853
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 residues
; TYPE: AMINO ACIDS
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: no
; ORIGINAL SOURCE:
; ORGANISM: Cattle pancreas
; FEATURE:
; OTHER INFORMATION: Peptide has two subunits. Second
; OTHER INFORMATION: subunit starts at position 22. Cys-Cys cross links
; OTHER INFORMATION: between positions 6 and 11, 7 and 28, and 20 and 40.
; OTHER INFORMATION: One or more of the Serine and/or Threonine residues
; OTHER INFORMATION: are phosphorylated as O-phosphates.
US-07-707-542E-3

Query Match 50.9%; Score 140; DB 1; Length 51;
Best Local Similarity 59.2%; Pred. No. 7.8e-13;
Matches 29; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

Qy 2 LVEQASTQASLYQIYNFNDVNFHLYGSHIREWLYLVAGERGFNDFPK 50
:111 11 1111: 1: 11 11 111: 1 111 11111: 111
Db 2 IVEQCTSTCSLYQLENYCNFVNHQHLGSHLVREALYLVCGRGFFFTPK 50

RESULT 4
US-07-707-542E-5
; Sequence 5, Application US/07707542E
; Patent No. 5242900
; GENERAL INFORMATION:
; APPLICANT: Albisser, A. Michael
; TITLE OF INVENTION: Treatment of Diabetes using
; TITLE OF INVENTION: Phosphorylated Insulin
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Queen's University
; STREET: PARTEQ Innovations
; CITY: Kingston
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 720K floppy
; COMPUTER: PC-AT compatible
; OPERATING SYSTEM: MSDOS 5.0
; SOFTWARE: Word for Windows 1.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07707,542E
; FILING DATE: 19910530

RESULT 5
US-07-707-542E-4
; Sequence 4, Application US/07707542E
; Patent No. 5242900
; GENERAL INFORMATION:
; APPLICANT: Alblisser, A. Michael
; TITLE OF INVENTION: Treatment of blabetes using
; TITLE OF INVENTION: Phosphorylated Insulin
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Queen's University
; STREET: PARTEQ Innovations
; CITY: Kingston
; COUNTRY: Canada
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 720K floppy
; COMPUTER: PC-AT compatible
; OPERATING SYSTEM: MSDOS 5.0
; SOFTWARE: Word for Windows 1.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07707,542E
; FILING DATE: 19910530
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hicks, Richard J.
; REGISTRATION NUMBER: 24665
; REFERENCE/DOCKET NUMBER: Q1341
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 545-2067
; TELEFAX: (613) 545-6853
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 residues
; TYPE: AMINO ACIDS
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/671,455

; FILING DATE: 03/19/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00303/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30
; TYPE: amino acid
; STRANDEDNESS: N/A
; TOPOLOGY: linear
US-08-212-696-2

Query Match 36.0%; Score 99; DB 1; Length 30;
Best Local Similarity 69.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

23 VNFHLYGSHIREWLIVAGRGFNFDPKT 51
|| || |||: | |||| |||||: |||
Db 2 VQHLGSHLVEALYLVCGERGFFYTPKT 30

RESULT 10
US-08-233-617-2
; Sequence 2, Application US/08233617
; Patent No. 5466666
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Sabel, Walter
; APPLICANT: Deil, Peter
; APPLICANT: Geisen, Karl
; TITLE OF INVENTION: Amorphous Monospherical Forms of Insulin
; TITLE OF INVENTION: Derivatives
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,617
; FILING DATE: 25-APR-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P 43 13 702.4
; FILING DATE: 27-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Carol P. Einaudi
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481.1374-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli
US-08-233-617-2

Query Match 36.0%; Score 99; DB 1; Length 30;
Best Local Similarity 69.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLIVAGRGFNFDPKT 51
|| || |||: | |||| |||||: |||
Db 2 VQHLGSHLVEALYLVCGERGFFYTPKT 30

RESULT 11
US-08-160-376A-2
; Sequence 2, Application US/08160376A
; Patent No. 5473049
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgen
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process For Obtaining Proinsulin
; TITLE OF INVENTION: Possessing Correctly Linked
; TITLE OF INVENTION: Cystine Bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth A. Genoni, Esq.
; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
; CITY: Somerville
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08876-1258
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 386
; OPERATING SYSTEM: WINDOWS 3.1
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/160,376A
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GE P 4240420.7
; FILING DATE: December 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara V. Maurer, Esq.
; REGISTRATION NUMBER: 31,287
; REFERENCE/DOCKET NUMBER: HOE 92/F 384
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 231-4079
; TELEFAX: (908) 231-2255
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 Amino Acids
; TYPE: Amino Acid
; TOPOLOGY: not relevant
US-08-160-376A-2

Query Match 36.0%; Score 99; DB 1; Length 30;
Best Local Similarity 69.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLIVAGRGFNFDPKT 51
|| || |||: | |||| |||||: |||
Db 2 VQHLGSHLVEALYLVCGERGFFYTPKT 30

RESULT 12
US-08-301-838-2
; Sequence 2, Application US/08301838
; Patent No. 5597796

```

; GENERAL INFORMATION:
; APPLICANT: Brange, Jens J.V.
; TITLE OF INVENTION: TRANSERMAL INSULIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 55977960 No. 5597796disk of No. 5597796th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/301,838
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/085,473
; FILING DATE: 30-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/751,836
; FILING DATE: 30-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0101/91
; FILING DATE: 22-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK92/00019
; FILING DATE: 22-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3447.210-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-9655
; TELEFAX: 212-867-0123
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 7
; OTHER INFORMATION: /note- "This disulfide bond is
; OTHER INFORMATION: between residue number 7 of this sequence and
; OTHER INFORMATION: residue number 7 of SEQ ID NO:1"
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 19
; OTHER INFORMATION: /note- "This disulfide bond is
; OTHER INFORMATION: between residue number 19 of this sequence and
; OTHER INFORMATION: residue number 20 of SEQ ID NO:1"
; US-08-301-838-2

Query Match 36.0%; Score 99; DB 1; Length 30;
Best Local Similarity 69.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYVAGERGFNFDPKT 51
   |||||:|||||:|||||:|||||:|||||
Db 2 VNHLCGSHLYEALYVCGERGFYTPKT 30

RESULT 13
US-08-389-487-5
; Sequence 5, Application US/08389487

Query Match 36.0%; Score 99; DB 1; Length 30;
Best Local Similarity 69.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYVAGERGFNFDPKT 51
   |||||:|||||:|||||:|||||:|||||
Db 2 VNHLCGSHLYEALYVCGERGFYTPKT 30

RESULT 14
US-08-508-564-7
; Sequence 7, Application US/08508664
; Patent No. 5840542
; GENERAL INFORMATION:
; APPLICANT: KANG, YUP
; APPLICANT: YOON, Ji-won
; TITLE OF INVENTION: METHOD FOR MANUFACTURE OF PROINSULIN
; TITLE OF INVENTION: WITH HIGH EXPORT YIELD
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: US
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; Patent No. 5663291
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgan
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process for Obtaining Insulin Having
; TITLE OF INVENTION: Correctly Linked Cystine Bridges
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,487
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481.1424-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-389-487-5

Query Match 36.0%; Score 99; DB 1; Length 30;
Best Local Similarity 69.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 23 VNFHLYGSHIREWLYVAGERGFNFDPKT 51
   |||||:|||||:|||||:|||||:|||||
Db 2 VNHLCGSHLYEALYVCGERGFYTPKT 30

RESULT 14
US-08-508-564-7
; Sequence 7, Application US/08508664
; Patent No. 5840542
; GENERAL INFORMATION:
; APPLICANT: KANG, YUP
; APPLICANT: YOON, Ji-won
; TITLE OF INVENTION: METHOD FOR MANUFACTURE OF PROINSULIN
; TITLE OF INVENTION: WITH HIGH EXPORT YIELD
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: US
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508.664
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, S. Peter
; REGISTRATION NUMBER: 25,351
; REFERENCE/DOCKET NUMBER: 0136/0B300
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: INSULIN B
;
US-08-508-664-7

```

```

Query Match 36.0%; Score 99; DB 2; Length 30;
Best Local Similarity 69.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

```

```

Qy 23 VNFHLYGSHIREWLYVAGERGFNFDPKT 51
   ||||| : ||||| : ||||| : ||||| : |||||
Db 2 VNHLCGSHLVEALYLVCGERGEFFYTPKT 30

```

```

RESULT 15
US-08-353-476-86
; Sequence 86, Application US/08353476
; Patent No. 5871902
; GENERAL INFORMATION:
; APPLICANT: Weininger, Susan
; APPLICANT: Weininger, Arthur M
; TITLE OF INVENTION: METHOD OF DETECTION OF DNA WITH A
; TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st St., Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353.476
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: GP-100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
;
US-08-353-476-86

```

```

Query Match 36.0%; Score 99; DB 2; Length 30;
Best Local Similarity 69.0%; Pred. No. 2.3e-07;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

```

```

Qy 23 VNFHLYGSHIREWLYVAGERGFNFDPKT 51
   ||||| : ||||| : ||||| : ||||| : |||||
Db 2 VNHLCGSHLVEALYLVCGERGEFFYTPKT 30

```

```

Search completed: August 23, 2002, 14:32:24
Job time: 137 sec

```


APPLICATION NUMBER: US 07/751,836
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 0101/91
FILING DATE: 22-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK92/00019
FILING DATE: 22-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3447.210-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 7
OTHER INFORMATION: /note= "This disulfide bond is
OTHER INFORMATION: between residue number 7 of this sequence and
OTHER INFORMATION: residue number 7 of SEQ ID NO:1"
FEATURE:
NAME/KEY: Disulfide-bond
LOCATION: 19
OTHER INFORMATION: /note= "This disulfide bond is
OTHER INFORMATION: between residue number 19 of this sequence and
OTHER INFORMATION: residue number 20 of SEQ ID NO:1"
US-08-301-838-2

Query Match 58.2%; Score 99; DB 1; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 VNFHLYGSHIREWLYLVAGERGFNDPKT 30
|| || ||| : |||| |||| : |||
Db 2 VNHLCGSHLVREALYLVCGERGFYTPKT 30

RESULT 5
US-08-389-487-5
Sequence 5, Application US/08389487
Patent No. 5663291
GENERAL INFORMATION:
APPLICANT: Obermeier, Rainer
APPLICANT: Gerl, Martin
APPLICANT: Ludwig, Jurgen
APPLICANT: Sabel, Walter
TITLE OF INVENTION: Process for Obtaining Insulin Having
TITLE OF INVENTION: Correctly Linked Cysteine Bridges
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk.
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/389,487
FILING DATE:

CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Einaudi, Carol P.
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481.1424-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-389-487-5

Query Match 58.2%; Score 99; DB 1; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 VNFHLYGSHIREWLYLVAGERGFNDPKT 30
|| || ||| : |||| |||| : |||
Db 2 VNHLCGSHLVREALYLVCGERGFYTPKT 30

RESULT 6
US-08-508-664-7
Sequence 7, Application US/08508664
Patent No. 5840542
GENERAL INFORMATION:
APPLICANT: Kang, Yup
APPLICANT: YOON, Ji-Won
TITLE OF INVENTION: METHOD FOR MANUFACTURE OF PROINSULIN
TITLE OF INVENTION: WITH HIGH EXPORT YIELD
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: US
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,664
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, S. Peter
REGISTRATION NUMBER: 25,351
REFERENCE/DOCKET NUMBER: 0136/0B300
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: INSULIN B
US-08-508-664-7

Query Match 58.2%; Score 99; DB 2; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
DB 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 7

US-08-353-476-86
; Sequence 86, Application US/08353476
; Patent No. 5871902
; GENERAL INFORMATION:
; APPLICANT: Weininger, Susan
; APPLICANT: Weininger, Arthur M
; TITLE OF INVENTION: METHOD OF DETECTION OF DNA WITH A
; TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st St., Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,476
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: GP-100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
US-08-353-476-86

Query Match 58.2%; Score 99; DB 2; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
DB 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 8

US-08-353-476-88
; Sequence 88, Application US/08353476
; Patent No. 5871902
; GENERAL INFORMATION:
; APPLICANT: Weininger, Susan
; APPLICANT: Weininger, Arthur M
; TITLE OF INVENTION: METHOD OF DETECTION OF DNA WITH A
; TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st St., Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,476
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: GP-100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
US-08-353-476-88

Query Match 58.2%; Score 99; DB 2; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNPHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
DB 2 VNQHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 9

US-08-484-219-6
; Sequence 6, Application US/08484219
; Patent No. 5911997
; GENERAL INFORMATION:
; APPLICANT: Schwabe, Christian
; APPLICANT: Unemori, Elaine
; TITLE OF INVENTION: RELAXIN-LIKE FACTOR AND METHODS AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,219
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 7842-040
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-484-219-6

Query Match 58.2%; Score 99; DB 2; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
Db 2 VNOHLCGSHLVEALYLVCGRGFFTPKT 30

RESULT 10
US-08-992-676-2
; Sequence 2, Application US/08992676
; Patent No. 5977297
; GENERAL INFORMATION:
; APPLICANT: OBERMEIER, Rainer
; APPLICANT: LUDWIG, Jurgen
; APPLICANT: SABEL, Walter
; TITLE OF INVENTION: A Process for Isolating Insulin Using
; TITLE OF INVENTION: High-Pressure Liquid Chromatography
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FINNEGAN HENDERSON FARABOW GARRETT & DUNNER,
; ADDRESSEE: L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,676
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jensen, Allen R.
; REGISTRATION NUMBER: 28,224
; REFERENCE/DOCKET NUMBER: 02481.1567-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli

Query Match 58.2%; Score 99; DB 2; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||

Db 2 VNOHLCGSHLVEALYLVCGRGFFTPKT 30
RESULT 11
US-09-134-836-2
; Sequence 2, Application US/09134836
; Patent No. 5986048
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; TITLE OF INVENTION: Insulin precursors having correctly bonded cystine bri
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,836
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDowell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli

Query Match 58.2%; Score 99; DB 2; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGRGFNFDPKT 30
|| || ||| : |||| |||| : |||
Db 2 VNOHLCGSHLVEALYLVCGRGFFTPKT 30

RESULT 12
US-08-750-391-4
; Sequence 4, Application US/08750391
; Patent No. 6015713
; GENERAL INFORMATION:
; APPLICANT: Wright Jr., James R.
; APPLICANT: Pohajdak, Bill
; TITLE OF INVENTION: Transgenic Fish in the Treatment of
; TITLE OF INVENTION: Diabetes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich
; STREET: 4365 Executive Drive, Suite 1600

CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,391
FILING DATE: 22-MAR-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/CA96/00171
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 90379
TELEPHONE: 619-677-1409
TELEFAX: 619-677-1465
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: HUMAN INSULIN B CHAIN
US-08-750-391-4

Query Match 58.2%; Score 99; DB 3; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNDPKT 30
DB 2 VNOHLCGSHLVEALYLVCGERGEFFTPKT 30

RESULT 13
US-09-255-668-1
Sequence 1, Application US/09255668
Patent No. 6197926
GENERAL INFORMATION:
APPLICANT: Gaur, Amitabh
APPLICANT: Ling, Nicholas
APPLICANT: Conlon, Paul J.
TITLE OF INVENTION: METHODS FOR TREATMENT OF DIABETES USING PEPTIDE
TITLE OF INVENTION: ANALOGUES OF INSULIN
FILE REFERENCE: 690068.448C1
CURRENT APPLICATION NUMBER: US/09/255,668
CURRENT FILING DATE: 1999-02-23
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1:
LENGTH: 30
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Human Insulin B Chain
US-09-255-668-1

Query Match 58.2%; Score 99; DB 4; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNDPKT 30
DB 2 VNOHLCGSHLVEALYLVCGERGEFFTPKT 30

RESULT 14
US-09-099-307-2
Sequence 2, Application US/09099307A
Patent No. 6221633
GENERAL INFORMATION:
APPLICANT: ERTL, JOHANN
APPLICANT: HABERMANN, PAUL
APPLICANT: GEISEN, KARL
APPLICANT: SEIPKE, GERHARD
TITLE OF INVENTION: NOVEL INSULIN DERIVATIVES HAVING A RAPID ONSET OF ACTION
FILE REFERENCE: 02481.1597-00000
CURRENT APPLICATION NUMBER: US/09/099,307A
CURRENT FILING DATE: 1998-06-18
EARLIER APPLICATION NUMBER: 19726167.1
EARLIER FILING DATE: 1997-06-20
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 30
TYPE: PRT
ORGANISM: Homo sapiens
US-09-099-307-2

Query Match 58.2%; Score 99; DB 4; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYLVAGERGFNDPKT 30
DB 2 VNOHLCGSHLVEALYLVCGERGEFFTPKT 30

RESULT 15
US-08-900-574-2
Sequence 2, Application US/08900574
Patent No. 6221837
GENERAL INFORMATION:
APPLICANT: Ertl, Johann
APPLICANT: Habermann, Paul
APPLICANT: Geisen, Karl
APPLICANT: Seipke, Gerhard
TITLE OF INVENTION: Insulin derivatives with increased zinc
TITLE OF INVENTION: binding
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett,
ADDRESS: & Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: District of Columbia
COUNTRY: U.S.A.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION NUMBER: US/08/900,574
FILING DATE: July 24, 1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: German Application No. 6221837 19630242.0
FILING DATE: July 26, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Carol P. Einaudi
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481.1499-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: Amino acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: Protein
ORIGINAL SOURCE: Escherichia coli
FEATURE:
NAME/KEY: Protein
LOCATION: 1..30
US-08-900-574-2

Query Match 58.2%; Score 99; DB 4; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 VNFHLYGSHIREWLYVAGERGFNFDPKT 30
||| ||| : |||| |||| : |||
Db 2 VNHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 16

US-09-386-303A-2
Sequence 2, Application US/09386303A
Patent No. 6380355
GENERAL INFORMATION:
APPLICANT: Rubroder, Franz-Josef
Keller, Reinhold

TITLE OF INVENTION: Improved process for obtaining
insulin precursors having correctly bonded cystine bridges

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
Dunner

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/09/386,303A

APPLICATION NUMBER: 09/134,836

FILING DATE: 31-Aug-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:

NAME: Leslie McDowell

REGISTRATION NUMBER: 34,872

REFERENCE/DOCKET NUMBER: 02481.1600-00000

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 30 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE: Escherichia coli

FEATURE:

NAME/KEY: Protein

LOCATION: 1..30

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-386-303A-2

Query Match 58.2%; Score 99; DB 4; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 VNFHLYGSHIREWLYVAGERGFNFDPKT 30
||| ||| : |||| |||| : |||
Db 2 VNHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 17

US-09-201-227A-20

Sequence 20, Application US/09201227A

Patent No. 6468770

GENERAL INFORMATION:

APPLICANT: Keyes, Linda N.

APPLICANT: Doberstein, Stephen K.

APPLICANT: Buchman, Andrew R.

TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS OF D. MELANOCASTER INSULIN-LIKE

TITLE OF INVENTION: GENES AND USES THEREOF

FILE REFERENCE: 7326-066

CURRENT APPLICATION NUMBER: US/09/201,227A

CURRENT FILING DATE: 1998-11-30

NUMBER OF SEQ ID NOS: 45

SOFTWARE: Patent In version 3.0

SEQ ID NO 20

LENGTH: 30

TYPE: PRT

ORGANISM: Homo sapiens

US-09-201-227A-20

Query Match 58.2%; Score 99; DB 4; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 VNFHLYGSHIREWLYVAGERGFNFDPKT 30
||| ||| : |||| |||| : |||
Db 2 VNHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 18

PCT-US94-04179-3

Sequence 3, Application PC/TUS9404179

GENERAL INFORMATION:

APPLICANT: Research Corporation Inc., Technologies

TITLE OF INVENTION: Methods and Compositions for Delaying or Preventing the

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee and Winner, P.C.

STREET: 5370 Manhattan Circle, #201

CITY: Boulder

STATE: Colorado

COUNTRY: USA

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/04179

FILING DATE: 15-APR-1994

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/048,979

FILING DATE: 16-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Ferber, Donna M.

REGISTRATION NUMBER: 33,878

REFERENCE/DOCKET NUMBER: 17-94PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 499-8080

TELEFAX: (303) 499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US94-04179-3

Query Match 58.2%; Score 99; DB 5; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYVAGRGFNFDPKT 30
|| || ||| : |||| ||||| : ||||
Db 2 VNHLCGSHLVEALYVCGRGFFTYPTKT 30

RESULT 19

5169865-5
Patent No. 5169865

APPLICANT: ANANTHANARAYANAN, V.S.
TITLE OF INVENTION: METHOD AND COMPOSITION FOR CALCIUM
BINDING TRANSLOCATION AND MEDIATING
NUMBER OF SEQUENCES: 12
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/323,421
FILING DATE: 14-MAR-1989
SEQ ID NO: 5:
LENGTH: 30

Query Match 58.2%; Score 99; DB 6; Length 30;
Best Local Similarity 69.0%; Pred. No. 4.7e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYVAGRGFNFDPKT 30
|| || ||| : |||| ||||| : ||||
Db 2 VNHLCGSHLVEALYVCGRGFFTYPTKT 30

RESULT 20

US-08-389-487-10
Sequence 10, Application US/08389487
Patent No. 5663291

GENERAL INFORMATION:
APPLICANT: Obermeier, Rainer
APPLICANT: Gerl, Martin
APPLICANT: Ludwig, Jurgen
APPLICANT: Sabel, Walter
TITLE OF INVENTION: Process for Obtaining Insulin Having
TITLE OF INVENTION: Correctly Linked Cysteine Bridges
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3315
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/389,487
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Jensen, Allen R.
REGISTRATION NUMBER: 28,224
REFERENCE/DOCKET NUMBER: 02481.1567-00000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-08-992-676-5

Query Match 58.2%; Score 99; DB 2; Length 31;
Best Local Similarity 69.0%; Pred. No. 4.9e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

NAME: Einaudi, Carol P.
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481.1424-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-389-487-10

Query Match 58.2%; Score 99; DB 1; Length 31;
Best Local Similarity 69.0%; Pred. No. 4.9e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLYVAGRGFNFDPKT 30
|| || ||| : |||| ||||| : ||||
Db 2 VNHLCGSHLVEALYVCGRGFFTYPTKT 30

RESULT 21

US-08-992-676-5
Sequence 5, Application US/08992676
Patent No. 5977297

GENERAL INFORMATION:
APPLICANT: OBERMEIER, Rainer
APPLICANT: LUDWIG, Jurgen
APPLICANT: SABEL, Walter
TITLE OF INVENTION: A Process for Isolating Insulin Using
TITLE OF INVENTION: High-Pressure Liquid Chromatography
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: FINNEGAN HENDERSON FARABOW GARRETT & DUNNER,
ADDRESSEE: L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,676
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Jensen, Allen R.
REGISTRATION NUMBER: 28,224
REFERENCE/DOCKET NUMBER: 02481.1567-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-08-992-676-5

Query Match 58.2%; Score 99; DB 2; Length 31;
Best Local Similarity 69.0%; Pred. No. 4.9e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLVLVAGRGFNFDPKT 30
||| ||| : ||| ||| : |||
Db 2 VNHLCGSHLVEALYLVCGRGFFFTPKT 30

RESULT 22
US-08-389-487-9
; Sequence 9, Application US/08389487
; Patent No. 5663291
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgin
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process for Obtaining Insulin Having
; TITLE OF INVENTION: Correctly Linked Cystine Bridges
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,487
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481.1424-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; STRANDEDNESS: single
; MOLECULE TYPE: peptide
; TOPOLOGY: linear

US-08-389-487-9
Query Match 58.2%; Score 99; DB 1; Length 32;
Best Local Similarity 69.0%; Pred. No. 5.1e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 VNFHLYGSHIREWLVLVAGRGFNFDPKT 30
||| ||| : ||| ||| : |||
Db 2 VNHLCGSHLVEALYLVCGRGFFFTPKT 30
RESULT 23
US-08-992-676-4
; Sequence 4, Application US/08992676
; Patent No. 5977297
; GENERAL INFORMATION:
; APPLICANT: OBERMEIER, Rainer
; APPLICANT: LUDWIG, Jurgin
; APPLICANT: SABEL, Walter
; TITLE OF INVENTION: A Process for Isolating Insulin Using
; TITLE OF INVENTION: High-Pressure Liquid Chromatography
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FINNEGAN HENDERSON FARABOW GARRETT & DUNNER,
; ADDRESS: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,836
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000

US-08-389-487-9
Query Match 58.2%; Score 99; DB 1; Length 32;
Best Local Similarity 69.0%; Pred. No. 5.1e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 VNFHLYGSHIREWLVLVAGRGFNFDPKT 30
||| ||| : ||| ||| : |||
Db 2 VNHLCGSHLVEALYLVCGRGFFFTPKT 30

RESULT 23
US-08-992-676-4
; Sequence 4, Application US/08992676
; Patent No. 5977297
; GENERAL INFORMATION:
; APPLICANT: OBERMEIER, Rainer
; APPLICANT: LUDWIG, Jurgin
; APPLICANT: SABEL, Walter
; TITLE OF INVENTION: A Process for Isolating Insulin Using
; TITLE OF INVENTION: High-Pressure Liquid Chromatography
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FINNEGAN HENDERSON FARABOW GARRETT & DUNNER,
; ADDRESS: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,836
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000

US-08-992-676-4
ADDRESSEE: L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,676
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jensen, Allen R.
REGISTRATION NUMBER: 28,224
REFERENCE/DOCKET NUMBER: 02481.1567-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-08-992-676-4

Query Match 58.2%; Score 99; DB 2; Length 32;
Best Local Similarity 69.0%; Pred. No. 5.1e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 VNFHLYGSHIREWLVLVAGRGFNFDPKT 30
||| ||| : ||| ||| : |||
Db 2 VNHLCGSHLVEALYLVCGRGFFFTPKT 30

RESULT 24
US-09-134-836-6
; Sequence 6, Application US/09134836
; Patent No. 5986048
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; TITLE OF INVENTION: Insulin precursors having correctly bonded cystine bri
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/134,836
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:

NAME/KEY: Protein

LOCATION: 1..32

US-09-134-836-6

Query Match 58.2%; Score 99; DB 2; Length 32;
Best Local Similarity 69.0%; Pred. No. 5.1e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLVLVAGERGDFNPKT 30
|| || |||: | |||| |||||: |||
DB 2 VNOHLCGSHLVEALYLVCGERGFFYTPKT 30

RESULT 25

US-09-386-303A-6
Sequence 6, Application US/09386303A
Patent No. 6380355

GENERAL INFORMATION:

APPLICANT: Rubroder, Franz-Josef
Keller, Reinhold

TITLE OF INVENTION: Improved process for obtaining

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/386.303A

FILING DATE: 31-Aug-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/134,836

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Leslie McDonnell

REGISTRATION NUMBER: 34,872

REFERENCE/DOCKET NUMBER: 02481.1600-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Escherichia coli

FEATURE:

NAME/KEY: Protein
LOCATION: 1..32
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-386-303A-6

Query Match 58.2%; Score 99; DB 4; Length 32;
Best Local Similarity 69.0%; Pred. No. 5.1e-08;
Matches 20; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 VNFHLYGSHIREWLVLVAGERGDFNPKT 30
|| || |||: | |||| |||||: |||
DB 2 VNOHLCGSHLVEALYLVCGERGFFYTPKT 30

Search completed: December 23, 2002, 07:17:53
Job time: 15.1176 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model
Run on: December 23, 2002, 07:14:17 ; Search time 25.5294 Seconds
(without alignments)
109.609 Million cell updates/sec

Title: US-09-574-443-7_COPY_1_21
Perfect score: 105
Sequence: 1 NLVEQASTSQSLYQIYNFDN 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: **BEST AVAILABLE COPY**

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	105	100.0	21	22	AAB48850	Mutant human insul
2	76	72.4	21	22	AAB48873	Mutant human insul
3	76	72.4	21	22	AAB48875	Mutant human insul
4	64	61.0	21	22	AAB48871	Mutant human insul
5	61	58.1	110	18	AAW22774	Mutant insulin pro
6	61	58.1	110	18	AAW22076	Human insulin chai
7	61	58.1	110	21	AAB36766	Mutant human insul
8	61	58.1	110	22	AAB35425	Human insulin muta
9	58	55.2	21	22	AAB48852	Mutant human insul
10	58	55.2	21	22	AAB48854	Mutant human insul

11	58	55.2	21	22	AAB48856	Mutant human insul
12	58	55.2	21	22	AAB48858	Mutant human insul
13	58	55.2	21	22	AAB48860	Mutant human insul
14	57	54.3	20	21	AAV99231	HLA class II bindi
15	55	52.4	21	13	AAR26469	Sequence of human-
16	55	52.4	57	21	AAV97990	RGD-proinsulin, co
17	54	51.4	21	7	AAP60910	Derivative of insu
18	54	51.4	21	7	AAP61327	Modified insulin A
19	54	51.4	21	10	AAP91590	Chain of superacti
20	54	51.4	21	13	AAR31331	Phosphorylated bo
21	54	51.4	21	22	AAB48862	Mutant human insul
22	53	50.5	21	6	AAP50251	Sequence of human
23	53	50.5	21	7	AAP60907	Derivative of insu
24	53	50.5	21	7	AAP61334	Modified insulin A
25	53	50.5	21	8	AAP71185	Sequence of A-chain
26	53	50.5	21	11	AAK04041	Amino acids 1-21 o
27	53	50.5	21	13	AAR27004	Phosphorylated por
28	53	50.5	21	13	AAR30649	Phosphorylated hum
29	53	50.5	21	13	AAR29998	Human insulin A-ch
30	53	50.5	21	14	AAR44453	Human insulin A-ch
31	53	50.5	21	15	AAP65880	Human insulin A-ch
32	53	50.5	21	15	AAR66217	Insulin (A1-A21).
33	53	50.5	21	15	AAR52572	Insulin A chain.
34	53	50.5	21	15	AAR53008	Human insulin A-ch
35	53	50.5	21	15	AAR68895	Human insulin A-ch
36	53	50.5	21	15	AAR57496	Insulin A chain an
37	53	50.5	21	16	AAR75412	Human insulin A ch
38	53	50.5	21	17	AAR94585	Insulin chain A an
39	53	50.5	21	17	AAR95999	Asymmetry sequence
40	53	50.5	21	17	AAR95872	Mono-acylated insu
41	53	50.5	21	17	AAR88747	Human insulin A-ch
42	53	50.5	21	19	AAW70340	Insulin analogue 1
43	53	50.5	21	19	AAW63754	Human insulin prot
44	53	50.5	21	19	AAW41318	Insulin analogue A
45	53	50.5	21	20	AAO17827	Human insulin 2 fr

ALIGNMENTS

RESULT 1
AAB48850
ID AAB48850 standard; peptide; 21 AA.
AC AAB48850;
XX
XX
DT 16-MAR-2001 (first entry)
DE Mutant human insulin (IA protein) cys1 A chain.
XX
XX
KW Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW disulphide bond; cysteine replacement; mutant; mutein.
XX
XX
OS Homo sapiens.
OS Synthetic.
XX
XX
PN WO200069901-A2.
XX
XX
PD 23-NOV-2000.
XX
XX
PF 19-MAY-2000; 2000WO-US13764.
XX
XX
PR 19-MAY-1999; 99US-0134930.
XX
XX
PA (XENC-) XENCOR INC.
XX
XX
PI Dahiyat BI;
XX
XX
WPI; 2001-025004/03.
XX
XX
PT Non-naturally occurring protein with insulin activity useful for
treating type 1 and type 2 diabetes, comprising amino acid

PT substitutions as compared to native human insulin and having enhanced
PT stability -
XX
PS Claim 10; Fig 3A; 95pp; English.
XX
CC The invention relates to novel non-naturally occurring mature human
CC insulin mutants, designated insulin activity (IA) proteins in the
CC specification (AAB48850-B48876), which have altered properties (e.g.,
CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
CC altered ability to oligomerise) compared to wild-type mature human
CC insulin (AAB48848, AAB48849) but which are still capable of binding to
CC an insulin receptor. The insulin mutants of the invention have less than
CC 98% sequence identity to wild-type human insulin, but have a 3D
CC structure which substantially corresponds to that of wild-type human
CC insulin. The invention also relates to recombinant nucleic acids encoding
CC the insulin mutants, expression vectors and host cells comprising mutant,
CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
CC and a pharmaceutical composition comprising an insulin mutant of the
CC invention. The mature human insulin mutants are useful for treating
CC insulin-responsive conditions and disorders of carbohydrate metabolism
CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
CC insulin mutants are used in gene therapy techniques for treating these
CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin
CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.
XX
SQ Sequence 21 AA;
Query Match 100.0%; Score 105; DB 22; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NLVEQASTSQASLYQIYNFDN 21
| | | | | | | | | | | | | | | | | | | | | |
Db 1 NLVEQASTSQASLYQIYNFDN 21
| | | | | | | | | | | | | | | | | | | | | |
RESULT 2
AAB48873
ID AAB48873 standard; peptide; 21 AA.
XX
AC AAB48873;
XX
DT 16-MAR-2001 (first entry)
XX
DE Mutant human insulin (IA protein) trz_07 A chain.
XX
KW Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2000069901-A2.
XX
PD 23-NOV-2000.
XX
PF 19-MAY-2000; 2000WO-US13764.
XX
PR 19-MAY-1999; 99US-0134930.
XX
PA (XENC-) XENCOR INC.
XX
PI Dahiyat BI;
XX
DR WPI; 2001-025004/03.
XX
PT Non-naturally occurring protein with insulin activity useful for
PT treating type 1 and type 2 diabetes, comprising amino acid
PT substitutions as compared to native human insulin and having enhanced

PT substitutions as compared to native human insulin and having enhanced
PT stability -
XX
PS Claim 10; Fig 5B; 95pp; English.
XX
CC The invention relates to novel non-naturally occurring mature human
CC insulin mutants, designated insulin activity (IA) proteins in the
CC specification (AAB48850-B48876), which have altered properties (e.g.,
CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
CC altered ability to oligomerise) compared to wild-type mature human
CC insulin (AAB48848, AAB48849) but which are still capable of binding to
CC an insulin receptor. The insulin mutants of the invention have less than
CC 98% sequence identity to wild-type human insulin, but have a 3D
CC structure which substantially corresponds to that of wild-type human
CC insulin. The invention also relates to recombinant nucleic acids encoding
CC the insulin mutants, expression vectors and host cells comprising mutant,
CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
CC and a pharmaceutical composition comprising an insulin mutant of the
CC invention. The mature human insulin mutants are useful for treating
CC insulin-responsive conditions and disorders of carbohydrate metabolism
CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
CC insulin mutants are used in gene therapy techniques for treating these
CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin
CC hexamers. Sequences AAB48871-B48876 represent the A and B chains of
CC insulin mutants which have improved stability.
XX
SQ Sequence 21 AA;
Query Match 72.4%; Score 76; DB 22; Length 21;
Best Local Similarity 71.4%; Pred. No. 7.5e-06;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 NLVEQASTSQASLYQIYNFDN 21
| : | | | | | | | | | | | | | | | | | | | | | |
Db 1 NIVEQCCTSOCSLYQIYNFCN 21
| : | | | | | | | | | | | | | | | | | | | | | |
RESULT 3
AAB48875
ID AAB48875 standard; peptide; 21 AA.
XX
AC AAB48875;
XX
DT 16-MAR-2001 (first entry)
XX
DE Mutant human insulin (IA protein) trz_08 A chain.
XX
KW Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2000069901-A2.
XX
PD 23-NOV-2000.
XX
PF 19-MAY-2000; 2000WO-US13764.
XX
PR 19-MAY-1999; 99US-0134930.
XX
PA (XENC-) XENCOR INC.
XX
PI Dahiyat BI;
XX
DR WPI; 2001-025004/03.
XX
PT Non-naturally occurring protein with insulin activity useful for
PT treating type 1 and type 2 diabetes, comprising amino acid
PT substitutions as compared to native human insulin and having enhanced
PT stability -

XX PS Claim 10; Fig 5C; 95pp; English.

XX CC The invention relates to novel non-naturally occurring mature human

CC insulin mutants, designated insulin activity (IA) proteins in the

CC specification (AAB48876), which have altered properties (e.g.,

CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,

CC altered ability to oligomerise) compared to wild-type mature human

CC insulin (AAB48848, AAB48849) but which are still capable of binding to

CC an insulin receptor. The insulin mutants of the invention have less than

CC 98% sequence identity to wild-type human insulin, but have a 3D

CC structure which substantially corresponds to that of wild-type human

CC insulin. The invention also relates to recombinant nucleic acids encoding

CC the insulin mutants, expression vectors and host cells comprising mutant,

CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,

CC and a pharmaceutical composition comprising an insulin mutant of the

CC invention. The mature human insulin mutants are useful for treating

CC insulin-responsive conditions and disorders of carbohydrate metabolism

CC such as type 1 or type 2 diabetes. The nucleic acids encoding these

CC insulin mutants are used in gene therapy techniques for treating these

CC conditions. The insulin mutants are more stable compared to wild-type

CC mature human insulin, or have an improved ability to form insulin

CC hexamers. Sequences AAB48871-B48876 represent the A and B chains of

CC insulin mutants which have improved stability.

XX CC

SQ Sequence 21 AA;

Query Match 72.4%; Score 76; DB 22; Length 21;

Best Local Similarity 71.4%; Pred. No. 7.5e-06;

Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 NLVEQASTSQASLYQIYNFDN 21

I:| | | | | | | | | | | | | | | |

DB 1 NIVEQCCTSQCSLYQYNYFCN 21

RESULT 4

AAB48871

ID AAB48871 standard; peptide; 21 AA.

XX AC AAB48871;

XX DT 16-MAR-2001 (first entry)

XX DE Mutant human insulin (IA protein) trz_06 A chain.

XX KW Human insulin; insulin activity protein; IA protein;

XX KW diabetes mellitus; type 1; type 2; stability;

XX KW mutant; mutin.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO200069901-A2.

XX PD 23-NOV-2000.

XX PF 19-MAY-2000; 2000WO-US13764.

XX PR 19-MAY-1999; 99US-0134930.

XX PA (XENC-) XENCOR INC.

XX PI Dahiyat BI;

XX WPI; 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for

PT treating type 1 and type 2 diabetes, comprising amino acid

PT substitutions as compared to native human insulin and having enhanced

PT stability -

XX Claim 10; Fig 5A; 95pp; English.

XX CC The invention relates to novel non-naturally occurring mature human

CC insulin mutants, designated insulin activity (IA) proteins in the

CC specification (AAB48850-B48876), which have altered properties (e.g.,

CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,

CC altered ability to oligomerise) compared to wild-type mature human

CC insulin (AAB48848, AAB48849) but which are still capable of binding to

CC an insulin receptor. The insulin mutants of the invention have less than

CC 98% sequence identity to wild-type human insulin, but have a 3D

CC structure which substantially corresponds to that of wild-type human

CC insulin. The invention also relates to recombinant nucleic acids encoding

CC the insulin mutants, expression vectors and host cells comprising mutant,

CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,

CC and a pharmaceutical composition comprising an insulin mutant of the

CC invention. The mature human insulin mutants are useful for treating

CC insulin-responsive conditions and disorders of carbohydrate metabolism

CC such as type 1 or type 2 diabetes. The nucleic acids encoding these

CC insulin mutants are used in gene therapy techniques for treating these

CC conditions. The insulin mutants are more stable compared to wild-type

CC mature human insulin, or have an improved ability to form insulin

CC hexamers. Sequences AAB48871-B48876 represent the A and B chains of

CC insulin mutants which have improved stability.

XX CC

SQ Sequence 21 AA;

Query Match 61.0%; Score 64; DB 22; Length 21;

Best Local Similarity 61.9%; Pred. No. 0.00075;

Matches 13; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 NLVEQASTSQASLYQIYNFDN 21

I:| | | | | | | | | | | | | | | |

DB 1 NIVEQCCTSQCSLYQYNYFCN 21

RESULT 5

AAW22774

ID AAW22774 standard; Protein; 110 AA.

XX AC AAW22774;

XX DT 25-MAR-1998 (first entry)

XX DE Mutant insulin protein which is incapable of forming disulphide bonds.

XX KW Insulin gene; regulation; insulin secretion; gene replacement;

XX KW type I diabetes; disulphide bond.

XX OS Synthetic.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Misc-difference 96

XX FT /label= C96S

XX FT /note= "wild type Cys substituted with Ser"

XX FT Misc-difference 109

XX FT /label= C109S

XX FT /note= "wild type Cys substituted with Ser"

XX PN WO9726334-A1.

XX PD 24-JUL-1997.

XX PF 17-JAN-1997; 97WO-US000760.

XX PR 19-JAN-1996; 96US-0589028.

XX PA (BETA-) BETAGENE INC.

XX PA (TEXA) UNIV. TEXAS SYSTEM.

XX PI Clark SA, Halban PA, Kruse F, Newgard CB, Normington KD;

XX PI Quade C, Thigpen AE;

XX WPI; 1997-393271/36.

DR N-PSDB; AAT75193.
 XX Improving production of secretory polypeptide(s), e.g. insulin - by
 PT blocking expression of native sequence and replacing with construct
 PT giving improved secretion, useful to treat, e.g. diabetes
 XX
 XX Example 5; Page 195; 279pp; English.
 XX
 CC The present sequence represents a mutant human insulin protein, which
 CC has 2 point mutations which change the cysteine at positions 96 and 109
 CC to serine. Both of these changes are in the A chain, and disrupt the 2
 CC disulphide bonds normally formed between A and B chains. The mutated
 CC insulin protein should be processed to human A, B, and C chains. Upon
 CC stimulated secretion, the 3 peptide chains would be released by the cell
 CC without the normal disulphide bonds. This mutated insulin can be used to
 CC prevent type I diabetes, by providing the mutated insulin gene, under
 CC the control of a promoter active in eukaryotic cells, to patients at
 CC risk. This method of treatment exemplifies a method for blocking the
 CC production of a native secreted polypeptide in a secretory host cell,
 CC and replacing it with exogenous polypeptide. This involves contacting the
 CC host cell with a polynucleotide encoding an exogenous polypeptide, under
 CC the control of a promoter active in eukaryotic cells and culturing the
 CC host cell under exogenous polypeptide expressing conditions. The method,
 CC which utilises the recombinant cells, can be used to prevent or treat
 CC type I diabetes. The method can be applied to improve or inhibit the
 CC secretion of any secretory peptide or protein, which may be advantageous
 CC in treating a disorder.
 XX
 SQ Sequence 110 AA;
 Query Match 58.1%; Score 61; DB 18; Length 110;
 Best Local Similarity 60.0%; Pred. No. 0.018;
 Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 2 LVEQASTSQASLYQIYNFDN 21
 :||| ||| ||||| :|
 Db 91 IVEQCSTSCSLYLENYSN 110
 RESULT 6
 ID AAW22076 standard; Protein; 110 AA.
 XX
 AC AAW22076;
 XX
 DT 03-FEB-1998 (first entry)
 XX
 DE Human insulin chain A disulphide mutant (C96S/C109S).
 XX
 KW Recombinant protein; expression; secretory cell line; insulin;
 KW diabetes; gene therapy; human.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX: WO9726321-A2.
 XX
 PD 24-JUL-1997.
 XX
 PF 17-JAN-1997; 97WO-US00761.
 XX
 PR 15-OCT-1996; 96US-0028427.
 PR 19-JAN-1996; 96US-0589028.
 XX
 PA (BETA-) BETAGENE INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Clark SA, Halban PA, Kruse F, McGarry D, Newgard CB;
 PI Normington KD, Quaade C, Thigpen AE;
 XX
 XX WPI; 1997-385326/35.
 DR N-PSDB; AAT75652.
 XX

PT Recombinant cell engineered to provide amylin to a mammal - useful
 PT to treat e.g. anorexia, anorexia, obesity, hypertension,
 XX osteoporosis etc.
 XX Example 4; Page 260-261; 336pp; English.
 XX
 CC This polypeptide comprises a disulphide mutant of human insulin A
 CC chain, in which the Cys-96 and Cys-109 residues of the wild-type
 CC sequence are substituted by Ser residues. These mutations disrupt
 CC the 2 disulphide bonds normally formed between the A and B chains.
 CC The mutant A chain is encoded by a DNA sequence. (see AAT75652)
 CC obtained by mutagenic PCR amplification of human insulin cDNA.
 CC The invention provides methods for production of heterologous
 CC polypeptides using recombinantly engineered cell lines. Also
 CC described are methods of engineering cells for high level
 CC expression, methods of large-scale heterologous protein production,
 CC and methods for treatment of disease in vivo using viral delivery
 CC systems and recombinant cell lines. A claimed method for
 CC preventing type I diabetes comprises providing a polynucleotide
 CC for a disulphide mutant of human insulin under control of a
 CC promoter active in eukaryotic cells to an individual at risk of
 CC type I diabetes.
 XX
 SQ Sequence 110 AA;
 Query Match 58.1%; Score 61; DB 18; Length 110;
 Best Local Similarity 60.0%; Pred. No. 0.018;
 Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 2 LVEQASTSQASLYQIYNFDN 21
 :||| ||| ||||| :|
 Db 91 IVEQCSTSCSLYLENYSN 110
 RESULT 7
 ID AAB26766 standard; Protein; 110 AA.
 XX
 AC AAB26766;
 XX
 DT 17-JAN-2001 (first entry)
 XX
 DE Mutant human insulin protein sequence.
 XX
 KW Amylin; production; secretory cell; blood glucose level regulation;
 KW diabetes mellitus; hypoglycaemia; osteoporosis; Paget's disease;
 KW hypercalcaemia; obesity; hypertension.
 XX
 OS Homo sapiens.
 XX
 PN US6110707-A.
 XX
 PD 29-AUG-2000.
 XX
 PF 17-JAN-1997; 97US-0784582.
 XX
 PR 11-OCT-1996; 96US-0028279.
 PR 19-JAN-1996; 96US-0589028.
 XX
 XX (TEXA) UNIV TEXAS SYSTEM.
 PA (BETA-) BETAGENE INC.
 XX
 XX Newgard CB, Halban P, Normington KD, Thigpen AE, Quaade C;
 PI Kruse F, McGarry D, Clark SA;
 XX
 XX WPI; 2000-586352/55.
 DR N-PSDB; AAC55717.
 XX
 XX Producing mammalian amylin, useful for regulating blood glucose and
 PT insulin levels, e.g. for treating diabetes mellitus or hypoglycaemia, by
 PT employing recombinantly engineered secretory cell lines -
 XX
 XX Example 4; Column 147-148; 136pp; English.
 PS

XX This invention relates to a method for producing mammalian amylin. The
 CC method relies on the use of a recombinantly engineered secretory cell
 CC line. The method comprises:
 CC (a) providing a starting secretory cell that has a regulated secretory
 CC pathway;
 CC (b) introducing, into the starting secretory cell, an amylin-encoding
 CC gene operatively linked to a first promoter;
 CC (c) selecting a secretory cell of (b) that exhibits increased production
 CC of biologically active amylin as compared to the starting secretory
 CC cell; and (d) culturing the selected secretory cell.
 CC Amylin is an insulin modulator, and the method results in antidiabetic,
 CC hypotensive and osteoprotic activity. The amylin produced are useful
 CC for regulating blood glucose levels, as well as in modulating the
 CC circulating levels of insulin in a mammal. The amylin produced maybe
 CC used in treating diabetes mellitus, hypoglycaemia, osteoporosis, Paget's
 CC disease, hypercalcaemia, obesity, hypertension, or any other disorder
 CC requiring amylin regulation. The invention includes cDNA and protein
 CC sequences (AAC55760 and AAB26771) representing human amylin. Sequences
 CC AAC55760 and AAB26771 are used in examples of the method of
 CC the invention for the production of mammalian amylin.

XX Sequence 110 AA;

Query Match 58.13; Score 61; DB 21; Length 110;
 Best Local Similarity 60.08; Pred. No. 0.018; 5; Indels 0; Gaps 0;
 Matches 12; Conservative 3; Mismatches 5;
 QY 2 LVEQASTSQASLQIYNFDN 21
 :||| ||| ||||| :|
 Db 91 IVEQSTSIQSGLQENYSN 110

RESULT 8
 AAB35425
 ID AAB35425 standard; Protein; 110 AA.

XX AAB35425;

XX 23-MAY-2001 (first entry)

XX Human insulin mutant SEQ ID NO: 4.

XX Secretory cell line; protein production; diabetes; disease treatment;
 KW hormone.

XX Homo sapiens.

XX US6194176-B1.

XX 27-FEB-2001.

XX 17-JAN-1997; 97US-0785271.

XX 19-JAN-1996; 96US-0589028.

XX (TEXA) UNIV TEXAS SYSTEM.

XX (BETA-) BETAGENE INC.

XX Newgard CB, Halban P, Normington KD, Clark SA, Thigpen AE;

XX Quade C, Kruse F;

XX WPI; 2001-217929/22.

XX Producing human insulin involves transforming secretory host cell with
 PT exogenous polynucleotide comprising gene encoding insulin, and
 PT culturing host cell

XX Example 5; Column 115-116; 107pp; English.

XX The present invention describes a method of producing human insulin,
 CC involving transforming a secretory host cell with an exogenous insulin
 CC coding sequence and culturing the cell in conditions suitable for insulin

CC expression. The invention also describes secretory cells useful as hosts.
 CC The method is also useful in the production of hormones and other
 CC proteins for use in disease treatment. The present sequence is a protein
 CC sequence used to demonstrate the method of the invention.

XX Sequence 110 AA;

Query Match 58.13; Score 61; DB 22; Length 110;
 Best Local Similarity 60.08; Pred. No. 0.018;
 Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLQIYNFDN 21

:||| ||| ||||| :|
 Db 91 IVEQSTSIQSGLQENYSN 110

RESULT 9
 AAB48852
 ID AAB48852 standard; peptide; 21 AA;

XX AAB48852;

XX 16-MAR-2001 (first entry)

XX Mutant human insulin (IA protein) cys77a A chain.

XX Human insulin; insulin activity protein; IA protein;

XX diabetes mellitus; type 1; type 2; stability;

XX disulphide bond; cysteine replacement; mutant; muten.

XX Homo sapiens.

XX Synthetic.

XX WO200069901-A2.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000WO-US13764.

XX 19-MAY-1999; 99US-0134930.

XX (XENC-) XENCOR INC.

XX Dahiyat BI;

XX WPI; 2001-025004/03.

XX Non-naturally occurring protein with insulin activity useful for
 PT treating type 1 and type 2 diabetes, comprising amino acid
 PT substitutions as compared to native human insulin and having enhanced
 PT stability

XX Claim 10; Fig 3B; 95pp; English.

XX The invention relates to novel non-naturally occurring mature human
 CC insulin mutants, designated insulin activity (IA) proteins in the
 CC specification (AAB48850-B48876), which have altered properties (e.g.,
 CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
 CC altered ability to oligomerise) compared to wild-type mature human
 CC insulin (AAB48848, AAB48849) but which are still capable of binding to
 CC an insulin receptor. The insulin mutants of the invention have less than
 CC 98% sequence identity to wild-type human insulin, but have a 3D
 CC structure which substantially corresponds to that of wild-type human
 CC insulin. The invention also relates to recombinant nucleic acids encoding
 CC the insulin mutants, expression vectors and host cells comprising mutant
 CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
 CC and a pharmaceutical composition comprising an insulin mutant of the
 CC invention. The mature human insulin mutants are useful for treating
 CC insulin-responsive conditions and disorders of carbohydrate metabolism
 CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
 CC insulin mutants are used in gene therapy techniques for treating these
 CC conditions. The insulin mutants are more stable compared to wild-type
 CC mature human insulin, or have an improved ability to form insulin

CC hexamers. Sequences AAB4850-B4863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.

SQ	Sequence	21 AA;
	Query Match	55.2%; Score 58; DB 22; Length 21;
	Best Local Similarity	60.0%; Pred. No. 0.0075;
	Matches 12; Conservative	3; Mismatches 5; Indels 0; Gaps 0;

```
QY      : 2 LVEQASTSQASLYQIYNFDN 21
          : ||| ||| ||||: | : |
Db      : 2 IVEOCSTSICSLYOLENYCN 21
```

RESULT 10
AAB48854
ID AAB48854 standard: peptide: 21 AA:

WPI; 2001-025004/03.

CC hexamers. Sequences ABA4850-B4863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.

Query Match 55.2%; Score 58; DB 22; Length 21;
Best Local Similarity 60.0%; Pred. No. 0.0075;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
QY      2 LVEQASTSQASLYQIYNFDN 21
      :||| ||| ||||: | : |
Db      2 IVEOCSTICSLYOLENYCN 21
```

RESULT 11
AAB48856
ID AAB48856 standard: peptide: 21 AA

Non-naturally occurring protein with insulin activity useful for treating type 1 and type 2 diabetes, comprising amino acid substitutions as compared to native human insulin and having enhanced stability -

CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.

XX
SQ Sequence 21 AA;

Query Match 55.2%; Score 58; DB 22; Length 21;
Best Local Similarity 60.0%; Pred. No. 0.0075;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 LVEQASTSOASLYQIYNFDN 21
:||| ||| ||||| :| :|
Db 2 IVEQCSISCSLYQENYCN 21

RESULT 12

AAB48858
ID AAB48858 standard; peptide; 21 AA.

XX
AC AAB48858;

XX
DT 16-MAR-2001 (first entry)

DE Mutant human insulin (IA protein) cysd+ A chain.

XX
DE Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW disulphide bond; cysteine replacement; mutant; mutein.

XX
OS Homo sapiens.

OS Synthetic.

XX
PN WO200069901-A2.

XX
PN 23-NOV-2000.

XX
PF 19-MAY-2000; 2000WO-US13764.

XX
PR 19-MAY-1999; 99US-0134930.

XX
PA (XENC-) XENCOR INC.

XX
PI Dahiyat BI;

XX
PS WPI; 2001-025004/03.

XX
PT Non-naturally occurring protein with insulin activity useful for
PT treating type 1 and type 2 diabetes, comprising amino acid
PT substitutions as compared to native human insulin and having enhanced
PT stability -

XX
PS Claim 10; Fig 3E; 95pp; English.

XX
CC The invention relates to novel non-naturally occurring mature human
CC insulin mutants, designated insulin activity (IA) proteins in the
CC specification (AAB48850-B48876), which have altered properties (e.g.,
CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
CC altered ability to oligomerise) compared to wild-type mature human
CC insulin (AAB48848, AAB48849) but which are still capable of binding to
CC an insulin receptor. The insulin mutants of the invention have less than
CC 98% sequence identity to wild-type human insulin, but have a 3D
CC structure which substantially corresponds to that of wild-type human
CC insulin. The invention also relates to recombinant nucleic acids encoding
CC the insulin mutants, expression vectors and host cells comprising mutant
CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
CC and a pharmaceutical composition comprising an insulin mutant of the
CC invention. The mature human insulin mutants are useful for treating
CC insulin-responsive conditions and disorders of carbohydrate metabolism
CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
CC insulin mutants are used in gene therapy techniques for treating these
CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin

CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.

XX
SQ Sequence 21 AA;

Query Match 55.2%; Score 58; DB 22; Length 21;
Best Local Similarity 60.0%; Pred. No. 0.0075;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 LVEQASTSOASLYQIYNFDN 21
:||| ||| ||||| :| :|
Db 2 IVEQCSISCSLYQENYCN 21

RESULT 13

AAB48860
ID AAB48860 standard; peptide; 21 AA.

XX
AC AAB48860;

XX
DT 16-MAR-2001 (first entry)

DE Mutant human insulin (IA protein) helix 24 A chain.

XX
DE Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW disulphide bond; cysteine replacement; mutant; mutein.

XX
OS Homo sapiens.

OS Synthetic.

XX
PN WO200069901-A2.

XX
PN 23-NOV-2000.

XX
PF 19-MAY-2000; 2000WO-US13764.

XX
PR 19-MAY-1999; 99US-0134930.

XX
PA (XENC-) XENCOR INC.

XX
PI Dahiyat BI;

XX
PS WPI; 2001-025004/03.

XX
PT Non-naturally occurring protein with insulin activity useful for
PT treating type 1 and type 2 diabetes, comprising amino acid
PT substitutions as compared to native human insulin and having enhanced
PT stability -

XX
PS Claim 10; Fig 3F; 95pp; English.

XX
CC The invention relates to novel non-naturally occurring mature human
CC insulin mutants, designated insulin activity (IA) proteins in the
CC specification (AAB48850-B48876), which have altered properties (e.g.,
CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
CC altered ability to oligomerise) compared to wild-type mature human
CC insulin (AAB48848, AAB48849) but which are still capable of binding to
CC an insulin receptor. The insulin mutants of the invention have less than
CC 98% sequence identity to wild-type human insulin, but have a 3D
CC structure which substantially corresponds to that of wild-type human
CC insulin. The invention also relates to recombinant nucleic acids encoding
CC the insulin mutants, expression vectors and host cells comprising mutant
CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
CC and a pharmaceutical composition comprising an insulin mutant of the
CC invention. The mature human insulin mutants are useful for treating
CC insulin-responsive conditions and disorders of carbohydrate metabolism
CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
CC insulin mutants are used in gene therapy techniques for treating these
CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin

CC hexamers. Sequences AAB4850-B4863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.

XX SQ Sequence 21 AA;

Query Match 55.2%; Score 58; DB 22; Length 21;
Best Local Similarity 60.0%; Pred. No. 0.0075;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Oy 2 LVEQASTSQASLYQIYNFDN 21
:|||||:|||||:|:
Db 2 IVEQSTSTCSLYQLENYCN 21

RESULT 14

AAIY9231
ID AAY99231 standard; Peptide; 20 AA.

XX AC AAY99231;

XX DT 07-AUG-2000 (first entry)

XX DE HLA class II binding antigen epitope peptide #420.

XX KW Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;
KW immune response; Chronic viral disease; cancer; autoimmune disease;
KW rheumatoid arthritis; multiple sclerosis; myasthenia gravis; AIDS;
KW allograft rejection; allergy; Lyme disease; hepatitis; prostate cancer;
KW glomerulonephritis; food hypersensitivity; malaria.

XX OS Homo sapiens.

XX FN WO9961916-A1.

XX PD 02-DEC-1999.

XX PF 28-MAY-1999; 99WO-US12066.

XX PR 29-MAY-1998; 98US-0087192.

XX PA (EPIM-) EPIMMUNE INC.

XX PI Sette A, Southwood S, Sidney J;

XX DR WPI; 2000-097143/08.

XX New compositions containing immunogenic peptide epitopes for various
XX HLA class II DR molecules useful for inducing helper T cell response

XX Claim 1; Page 47; 60pp; English.

XX The present invention relates to a new pharmaceutical composition
XX comprising a unit dose form of a peptide, or analogue, comprising an
XX epitope selected from those represented by peptides AAY98812-Y99339
XX which are derived from various antigens for various human leucocyte
XX antigen class DR molecules, representative of the world wide population.
XX The peptide/analogue binds to an HLA class II molecule at an IC-50 of
XX less than or equal to 1,000 nM. The pharmaceutical can be used to induce
XX a helper T cell response. The pharmaceutical focuses the immune response
XX towards selected determinants and could therefore be used in cases of
XX chronic viral diseases and cancer. Examples of diseases that can be
XX treated using the peptide containing pharmaceutical include autoimmune
XX diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia
XX gravis), allograft rejection, allergies, Lyme disease, hepatitis,
XX post-streptococcal endocarditis or glomerulonephritis and food
XX hypersensitivities. The peptide epitopes can be used to enhance immune
XX responses against other immunogens administered with the peptides.
XX Diseases which can be treated using immunogenic mixtures include
XX prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma,
XX cervical carcinoma, lymphoma, and condyloma acuminatum. The peptides may
XX also be used to make monoclonal antibodies useful as potential diagnostic

CC or therapeutic agents. The peptides may also be useful as diagnostic
CC reagents, for example, to determine the susceptibility of an individual
CC to a treatment regimen. Also, the peptides may be used to predict which
CC individuals will be at substantial risk of developing chronic infection.
CC The selection of appropriate T and B cell epitopes should allow the
CC development of epitope based vaccines particularly towards conserved
CC epitopes of pathogens which are characterized by high sequence
CC variability such as HIV, HCV and Malaria.

XX SQ Sequence 20 AA;

Query Match 54.3%; Score 57; DB 21; Length 20;
Best Local Similarity 61.1%; Pred. No. 0.01;
Matches 11; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 2 LVEQASTSQASLYQIYNF 19

:|||||:|||||:|:

Db 2 IVEQSTSTCSLYQLENY 19

RESULT 15

AAIY9231

ID AAR26469 standard; Protein; 21 AA.

XX AC AAR26469;

XX DT 28-JAN-1993 (first entry)

XX DE Sequence of human insulin A-chain.

XX KW Serine protease; proteolytic enzyme; plastein reaction; insulin;
KW SP 446 protease.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Disulfide-bond 6..11

FT Disulfide-bond 7

FT /note= "bonded to AA-7 of B-chain"

FT Disulfide-bond 20

FT /note= "bonded to AA-19 of B-chain"

FT Cleavage-site 4..5

FT /note= "see CC"

FT Cleavage-site 17..18

FT /note= "see CC"

XX PN WO9213964-A.

XX PD 20-AUG-1992.

XX PF 06-FEB-1992; 92WO-DK00036.

XX PR 06-FEB-1991; 91DK-0000199.

XX (NOVO) NOVO-NORDISK AS.

XX Budtz P, Nielsen P;

XX WPI; 1992-300049/36.

XX Increasing viscosity of protein prep(s). - by incubating with
XX serine protease specific for glutamic acid and aspartic acid
XX residues

XX Example; Fig 3; 38pp; English.

XX The serine protease of the invention has the following
XX characteristics: (a) it is a serine protease specific for glutamic
XX acid and aspartic acid residues; (b) it has a specific activity of
XX >25 CPU/g of enzyme protein; (c) it has an apparent mol. wt. of
XX 23,600; (d) it is inhibited by DPP but not by PMSF, and (e) it
XX exhibits 75% or more of its maximizing activity in the pH range 6.5-
XX 10. In partic. it is AAR26467. It was previously characterized in


```
XX AAP61327;
AC
XX
XX
DT 23-OCT-1991 (first entry)
XX
DE Modified insulin A-chain.
XX
XX Diabetes.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Disulfide-bond 6..11
FT Cross-links 20
FT /label- Disulfide link to B19
XX
XX W08605497-A.
PN
XX
XX 25-SEP-1986.
PD
XX
XX 14-MAR-1986; 86WO-DK00023.
PF
XX
XX 14-MAR-1986; 86WO-DK00023.
PR
XX 15-MAR-1985; 85DK-0001197.
XX
XX (NOGE-) NORDISK GENTOFTE AS.
PA
XX
XX Balschmidt P, Hansen FB;
PI
XX
XX WPI; 1986-264939/40.
DR
XX
XX New insulin derivs. with uncharged side chain at A4, A17, B13 or
PT B21 - useful as prolonged acting forms of insulin.
PT
XX
XX Claim 7; Page 15; 18pp; English.
PS
XX
XX The novel insulin A-chain derivative has uncharged Gln replacing
CC Glu17. The resulting peptide has prolonged action, reducing the
CC number of injections required daily. The product may also be used in
CC a preparation with a fast-acting insulin.
XX
XX Sequence 21 AA;
SQ
Query Match 51.4%; Score 54; DB 7; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.035;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 2 IVEQASTSOASLYQIYNFDN 21
:|||||
DB 2 IVEQCCTSLCSLYQLQNYCN 21
:|||||
RESULT 19
AAP91590
ID AAP91590 standard; protein; 21 AA.
XX
XX
AC AAP91590;
XX
XX 02-JUL-1990 (first entry)
DT
XX
XX Chain of superactive insulin analogue.
DE
XX
XX Insulin; diabetes; insulin analogue.
KW
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Disulfide-bond 6
FT /note="Bonds to 11th residue following (Cys)".
FT Disulfide-bond 7
FT /note="Bonds to 7th residue of AAP91591 (Cys)".
FT Disulfide-bond 20
FT /note="Bonds to 19th residue of AAP91591 (Cys)".
FT
```

```
XX W0900580-A.
PN
XX
XX 26-JAN-1989.
PD
XX
XX 07-JUL-1988; 88WO-US02289.
PF
XX
XX 17-JUL-1987; 87US-0074558.
PR
XX
XX (MOUN-) MOUNT SINAI SCHOOL.
PA
XX
XX Katsoyannis PG, Schwartz GP;
PI
XX
XX WPI; 1989-054068/07.
DR
XX
XX New human insulin analogues for the treatment of diabetes -
PT have B-chain modified by substitution of histidine (10) and opt.
PT also carboxyl-terminus penta-peptide deletion.
XX
XX Claim 1; Page 38; 50pp; English.
PS
XX
XX Chain forms half superactive insulin analogue, joined to second chain by
CC two disulfide bridges.
CC
XX Second chain is AAP91591.
XX
XX Sequence 21 AA;
SQ
Query Match 51.4%; Score 54; DB 10; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.035;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 2 IVEQASTSOASLYQIYNFDN 21
:|||||
DB 2 IVEQCCTSLCSLYQLQNYCN 21
:|||||
RESULT 20
AAR31331
ID AAR31331 standard; protein; 21 AA.
XX
XX
AC AAR31331;
XX
XX 20-MAY-1998 (first entry)
DT
XX
XX Phosphorylated bovine insulin A-chain.
DE
XX
XX Diabetes mellitus; N-phosphorylated amino acid; blood glucose;
KW control; insulin-dependent diabetes; isoelectric point.
XX
XX Bos taurus.
OS
XX
XX Key Location/Qualifiers
FH modified_site 1
FT /note="opt. phosphorylated"
FT modified_site 8
FT /note="opt. phosphorylated"
FT modified_site 9
FT /note="opt. phosphorylated"
FT modified_site 12
FT /note="opt. phosphorylated"
FT modified_site 14
FT /note="opt. phosphorylated"
FT modified_site 18
FT /note="opt. phosphorylated"
FT modified_site 19
FT /note="opt. phosphorylated"
FT modified_site 21
FT /note="opt. phosphorylated"
XX
XX W09214754-A.
PN
XX
XX 03-SEP-1992.
PD
XX
```

PF 26-FEB-1992; 92WO-CA00082.
XX
PR 26-FEB-1991; 91GB-0004037.
XX
PA (HOSP-) HOSPITAL SICK CHILDREN LOYAL TRUE BLUE.
XX
PI Loughheed WD;
XX
DR WPI; 1992-316123/38.
XX
PT Phosphorylation of peptide(s) with phosphorus oxychloride - esp.
PT for prodn. of phosphorylated insulin for diabetes treatment
XX
XX Claim 14-17; Page 10; 46pp; English.
XX
CC Phosphorylated insulin is claimed for treatment of diabetes in
CC which at least one of the tyrosine residues is phosphorylated. The
CC other residues indicated in the features table can also be
CC phosphorylated. Phosphorylation is effected by reacting an aqueous
CC solution of insulin with phosphorus oxychloride. The phosphorylated
CC insulin is purified and only those portions which have
CC substantially reduced isoelectric points are isolated. Unmodified
CC insulin is removed to prevent negation of the benefit obtained from
CC the 'gentler pharmacokinetic properties' of the phosphorylated
CC insulin.
CC See also AAR31332 for phosphorylated B-chain.
XX
SQ Sequence 21 AA;

Query Match 51.4%; Score 54; DB 13; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.035;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:111 11 1111:1:
Db 2 IVEQCCTSVCSLYQLENYCN 21

RESULT 21
AAB48862
ID AAB48862 standard; peptide; 21 AA.
XX
AC AAB48862;
XX
DT 16-MAR-2001 (first entry)
XX
DE Mutant human insulin (IA protein) cys-4 A chain.
XX
KW Human insulin; insulin activity protein; IA protein;
KW diabetes mellitus; type 1; type 2; stability;
KW disulphide bond; cysteine replacement; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
XX
XX WO200069901-A2.
XX
XX 23-NOV-2000.
XX
XX 19-MAY-2000; 2000WO-US13764.
XX
XX 19-MAY-1999; 99US-0134930.
XX
XX (XENC-) XENCOR INC.
XX
XX Dahiyat BI;
XX
XX WPI; 2001-025004/03.
XX
XX Non-naturally occurring protein with insulin activity useful
XX for treating type 1 and type 2 diabetes, comprising amino acid
XX substitutions as compared to native human insulin and having enhanced
XX stability.

XX PS Claim 10; Fig 3G; 95pp; English.
XX
CC The invention relates to novel non-naturally occurring mature human
CC insulin mutants, designated insulin activity (IA) proteins in the
CC specification (AAB48850-B48876), which have altered properties (e.g.,
CC enhanced oxidative, alkaline or thermal stability, prolonged shelf-life,
CC altered ability to oligomerise) compared to wild-type mature human
CC insulin (AAB48848, AAB48849) but which are still capable of binding to
CC an insulin receptor. The insulin mutants of the invention have less than
CC 98% sequence identity to wild-type human insulin, but have a 3D
CC structure which substantially corresponds to that of wild-type human
CC insulin. The invention also relates to recombinant nucleic acids encoding
CC the insulin mutants, expression vectors and host cells comprising mutant,
CC insulin-encoding DNA, the recombinant preparation of an insulin mutant,
CC and a pharmaceutical composition comprising an insulin mutant of the
CC invention. The mature human insulin mutants are useful for treating
CC insulin-responsive conditions and disorders of carbohydrate metabolism
CC such as type 1 or type 2 diabetes. The nucleic acids encoding the
CC insulin mutants are used in gene therapy techniques for treating these
CC conditions. The insulin mutants are more stable compared to wild-type
CC mature human insulin, or have an improved ability to form insulin
CC hexamers. Sequences AAB48850-B48863 represent the A and B chains of
CC insulin mutants in which one or more cysteine residue has been
CC replaced, thus preventing the formation of at least one disulphide bond
CC and improving stability and activity.
XX
SQ Sequence 21 AA;

Query Match 51.4%; Score 54; DB 22; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.035;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:111 11 1111:1:
Db 2 IVEQCCTSVCSLYQLENYCN 21

RESULT 22
AAP50251
ID AAP50251 standard; protein; 21 AA.
XX
AC AAP50251;
XX
DT 09-MAR-1992 (first entry)
XX
DE Sequence of human insulin Chain A deriv.
XX
KW Diabetes; therapy; hormone.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Modified-site 1
FT /label= H-Gly
FT Disulfide-bond 6..11
FT Disulfide-bond 7
FT /note= "bonded to Cys7 of Chain B (AAP51010)"
FT Disulfide-bond 20
FT /note= "bonded to Cys19 of Chain B (AAP51010)"
FT Modified-site 21
FT /label= Asn-OH
XX
XX EP132770-A.
XX
XX 13-FEB-1985.
XX
XX 18-JUL-1984; 84EP-0108442.
XX
XX 22-JUL-1983; 83DE-3326472.
XX
XX (FARH) HOECHST AG.
XX

PI Grau U;
XX WPI; 1985-039407/07.
XX
XX New insulin derivs. with modified B chain C terminus - having
XX long lasting action without requiring slow release additive
XX
XX Claim 1; Page 1; 43pp; German.
XX
XX The inventors claim new insulin derivs. with modified B chain C
XX terminus. The claimed insulin derivs. provide a closer match to the
XX natural hormonal equilibrium than known forms of insulin; are no
XX more immunogenic; can be administered in less strongly acid solns.
XX and (because of their low solubility at physiological pH) provide a
XX long-lasting action without delayed-release additives.
XX
XX Sequence 21 AA;
XX
XX Query Match 50.5%; Score 53; DB 6; Length 21;
XX Best Local Similarity 55.0%; Pred. No. 0.051;
XX Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
XX
XX QY 2 LVEQASTQASLYQIYNFDN 21
XX :||| || ||||: |:
XX DB 2 IVEQCTSCSYQLENYCN 21
XX
XX RESULT 23
XX AAP60907
XX ID AAP60907 standard; peptide; 21 AA.
XX AC AAP60907;
XX XX
XX 26-JUN-1991 (first entry)
XX
XX Derivative of insulin A-chain.
XX
XX Insulin; hormone.
XX
XX Key Location/Qualifiers
XX Misc-difference 4 /label= Glu or neutral AA
XX Disulfide-bond 7 /label= links with Cys7 of Insulin B-chain
XX Misc-difference 17 /label= Glu or neutral AA
XX Disulfide-bond 20 /label= links with Cys19 of Insulin B-chain
XX
XX EP194864-A.
XX
XX 17-SEP-1986.
XX
XX 11-MAR-1986; 86EP-0301755.
XX
XX 12-MAR-1985; 85DK-0001135.
XX
XX (NOVO) NOVO INDUSTRI A/S.
XX
XX Markussen J;
XX WPI; 1986-246980/38.
XX
XX New insulin derivs. with prolonged activity - prep'd. by
XX transpeptidating porcine insulin using trypsin catalyst.
XX
XX Claim 1; Page 29; 33pp; English.
XX
XX The modified A-chain of insulin is linked to the B-chain by
XX disulfide bonds between Cys-7 (A-chain) and Cys-7 (B-chain), and
XX between Cys-20 (A-chain) and Cys-19 (B-chain). The insulin
XX derivatives have prolonged insulin action, the degree of which
XX may be enhanced and controlled by Zn²⁺. The derivatives may

CC be prepared by transpeptidating pig insulin using trypsin.
CC See also AAP60908.
XX
XX Sequence 21 AA;
XX
XX Query Match 50.5%; Score 53; DB 7; Length 21;
XX Best Local Similarity 55.0%; Pred. No. 0.051;
XX Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
XX
XX QY 2 LVEQASTQASLYQIYNFDN 21
XX :||| || ||||: |:
XX DB 2 IVEQCTSCSYQLENYCN 21
XX
XX RESULT 24
XX AAP61334
XX ID AAP61334 standard; protein; 21 AA.
XX AC AAP61334;
XX XX
XX 23-OCT-1991 (first entry)
XX
XX Modified insulin A-chain.
XX
XX Diabetes.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Modifie-site 4 /label= Ala, OTHER
XX Disulfide-bond 6..11 /note= "diamide or tetramide"
XX Modified-site 17
XX FT /label= OTHER
XX FT /note= "dimethyl ester, amide, diamide or tetramide"
XX FT Cross-links 20
XX FT /label= Disulfide link to B19
XX
XX WO8605496-A.
XX
XX 25-SEP-1986.
XX
XX 14-MAR-1986; 86WO-DK00022.
XX
XX 15-MAR-1985; 85DK-0001197.
XX 14-MAR-1986; 86WO-DK00023.
XX 14-MAR-1986; 86WO-DK00022.
XX 14-NOV-1986; 86DK-0005456.
XX 14-NOV-1986; 86DK-0005457.
XX
XX (NOGE-) NORDISK GENTOFTE AS.
XX (BALS/) BALSCHMIDT P.
XX (NOVO) NOVO-NORDISK A/S.
XX
XX Balschmidt P, Hansen F;
XX
XX WPI; 1986-264938/40.
XX
XX Novel insulin derivs. with protracted action - has B30 carboxyl
XX gp. blocked with uncharged gp.
XX
XX Claim 4; Page 19; 18pp; English.
XX
XX Modified insulin A-chain may have one or more protected groups which act
XX to protract the effective life of the hormone, varying in degree
XX according to blocking groups.
XX
XX Sequence 21 AA;
XX
XX Query Match 50.5%; Score 53; DB 7; Length 21;
XX Best Local Similarity 55.0%; Pred. No. 0.051;
XX Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```

Search completed: December 23, 2002, 07:17:22
Job time : 27.5294 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 23, 2002, 07:14:22 ; Search time 9.88235 Seconds
(without alignments)
62.524 Million cell updates/sec

Title: US-09-574-443-7_COPY_1_21

Perfect score: 105
Sequence: 1 NLVEQASTSQASLIQIYNFDN 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_5/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_5/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_5/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_5/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_5/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_5/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	58.1	110	3	US-08-589-028-4 Sequence 4, Appl1
2	61	58.1	110	3	US-08-784-582-4 Sequence 4, Appl1
3	61	58.1	110	4	US-08-785-271-4 Sequence 4, Appl1
4	53	50.5	21	1	US-08-212-696-1 Sequence 1, Appl1
5	53	50.5	21	1	US-08-158-245-1 Sequence 1, Appl1
6	53	50.5	21	1	US-08-233-617-1 Sequence 1, Appl1
7	53	50.5	21	1	US-08-160-376A-1 Sequence 1, Appl1
8	53	50.5	21	1	US-08-304-070-1 Sequence 1, Appl1
9	53	50.5	21	1	US-08-285-661-1 Sequence 1, Appl1
10	53	50.5	21	1	US-08-301-838-1 Sequence 1, Appl1
11	53	50.5	21	1	US-08-389-487-4 Sequence 4, Appl1
12	53	50.5	21	1	US-08-507-124-2 Sequence 2, Appl1
13	53	50.5	21	1	US-08-342-931-1 Sequence 1, Appl1
14	53	50.5	21	2	US-08-508-664-9 Sequence 9, Appl1
15	53	50.5	21	2	US-08-353-476-85 Sequence 85, Appl1
16	53	50.5	21	2	US-08-484-219-1 Sequence 1, Appl1
17	53	50.5	21	2	US-08-979-587-1 Sequence 1, Appl1
18	53	50.5	21	2	US-08-992-676-1 Sequence 1, Appl1
19	53	50.5	21	2	US-09-134-836-1 Sequence 1, Appl1
20	53	50.5	21	3	US-08-967-867-2 Sequence 2, Appl1
21	53	50.5	21	3	US-08-750-391-3 Sequence 3, Appl1
22	53	50.5	21	4	US-09-099-307-1 Sequence 1, Appl1
23	53	50.5	21	4	US-08-900-574-1 Sequence 1, Appl1
24	53	50.5	21	4	US-08-552-749-1 Sequence 1, Appl1
25	53	50.5	21	4	US-09-386-303A-1 Sequence 1, Appl1
26	53	50.5	21	4	US-09-201-227A-21 Sequence 21, Appl1
27	53	50.5	21	6	5164366-3 Patent No. 5164366

28 53 50.5 21 6 5169865-4 Patent No. 5169865
29 53 50.5 21 6 5514646-50 Patent No. 5514646
30 53 50.5 51 1 US-07-707-542E-1 Sequence 1, Appl1
31 53 50.5 51 1 US-07-707-542E-2 Sequence 2, Appl1
32 53 50.5 51 4 US-09-477-924-3 Sequence 3, Appl1
33 53 50.5 53 1 US-08-233-617-4 Sequence 4, Appl1
34 53 50.5 53 4 US-09-261-853-2 Sequence 2, Appl1
35 53 50.5 53 4 US-08-981-988A-42 Sequence 42, Appl1
36 53 50.5 56 1 US-08-160-376A-7 Sequence 7, Appl1
37 53 50.5 56 1 US-08-389-487-11 Sequence 11, Appl1
38 53 50.5 57 1 US-08-030-731A-44 Sequence 44, Appl1
39 53 50.5 63 1 US-08-160-376A-6 Sequence 6, Appl1
40 53 50.5 65 1 US-08-468-674B-71 Sequence 71, Appl1
41 53 50.5 65 1 US-08-780-571-71 Sequence 71, Appl1
42 53 50.5 66 1 US-08-291-960B-5 Sequence 5, Appl1
43 53 50.5 67 4 US-08-981-988A-1 Sequence 1, Appl1
44 53 50.5 67 4 US-08-981-988A-2 Sequence 2, Appl1
45 53 50.5 67 4 US-08-981-988A-5 Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-08-589-028-4
; Sequence 4, Application US/08589028
; Patent No. 6087129
; GENERAL INFORMATION:
; APPLICANT: Newgard, Christopher B.
; APPLICANT: Halban, Philippe
; APPLICANT: No. 6087129mington, Karl D.
; APPLICANT: Clark, Samuel A.
; APPLICANT: Thigpen, Anice E.
; APPLICANT: Quade, Christian
; APPLICANT: Kruse, Fred
; TITLE OF INVENTION: Recombinant Expression of Proteins From
; TITLE OF INVENTION: Secretory Cell Lines
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/589,028
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 47,642
; REFERENCE/DOCKET NUMBER: UTSD:426\HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 110 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-589-028-4

Query Match 58.1%; Score 61; DB 3; Length 110;
Best Local Similarity 60.0%; Pred. No. 0.003;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
:||||| 1111 1111: 1: 1
Db 91 IVEQCSTICSLYQLENYSN 110

RESULT 2
US-08-784-582-4
; Sequence 4, Application US/08784582
; Patent No. 6110707
; GENERAL INFORMATION:
; APPLICANT: Newgard, Christopher B.
; APPLICANT: Halban, Philippe A.
; APPLICANT: No. 6110707/mington, Karl D.
; APPLICANT: Clark, Samuel A.
; APPLICANT: Thigpen, Anice E.
; APPLICANT: Quaade, Christian
; APPLICANT: Kruse, Fred
; APPLICANT: McGarry, Dennis
; TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM
; SECRETORY CELL LINES
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,427
; FILING DATE: 15-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/589,028
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSD:514
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 110 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-784-582-4

Query Match 58.1%; Score 61; DB 3; Length 110;
Best Local Similarity 60.0%; Pred. No. 0.003;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
:||||| 1111 1111: 1: 1
Db 91 IVEQCSTICSLYQLENYSN 110

RESULT 3
US-08-785-271-4
; Sequence 4, Application US/08785271
; Patent No. 6194176
; GENERAL INFORMATION:
; APPLICANT: Newgard, Christopher B.

; APPLICANT: Halban, Philippe A.
; APPLICANT: No. 6194176/mington, Karl D.
; APPLICANT: Clark, Samuel A.
; APPLICANT: Thigpen, Anice E.
; APPLICANT: Quaade, Christian
; APPLICANT: Kruse, Fred
; TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM
; SECRETORY CELL LINES
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,271
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/589,028
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSD:513
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 110 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-785-271-4

Query Match 58.1%; Score 61; DB 4; Length 110;
Best Local Similarity 60.0%; Pred. No. 0.003;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
:||||| 1111 1111: 1: 1
Db 91 IVEQCSTICSLYQLENYSN 110

RESULT 4
US-08-212-696-1
; Sequence 1, Application US/08212696
; Patent No. 5422339
; GENERAL INFORMATION:
; APPLICANT: George S. Eisenbarth et al.
; TITLE OF INVENTION: PEPTIDES HAVING INSULIN
; TITLE OF INVENTION: AUTOANTIBODY BUT NOT
; TITLE OF INVENTION: INSULIN RECEPTOR BINDING
; TITLE OF INVENTION: CAPACITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX

;; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
;; SOFTWARE: WordPerfect (Version 5.0)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/212,696
;; FILING DATE:
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/671,455
;; FILING DATE: 03/19/91
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Freeman, John W.
;; REGISTRATION NUMBER: 29,066
;; REFERENCE/DOCKET NUMBER: 00303/009001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 542-5070
;; TELEFAX: (617) 542-8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 21
;; TYPE: amino acid
;; STRANDEDNESS: N/A
;; TOPOLOGY: linear
US-08-212-696-1

Query Match 50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087; 6; Indels 0; Gaps 0;
Matches 11; Conservative 3; Mismatches 6;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||: 1: 1
DB 2 IVEQCTSCISLYQLENYCN 21

RESULT 5
US-08-158-245-1
;; Sequence 1, Application US/08158245
;; Patent No. 5442043
;; GENERAL INFORMATION:
;; APPLICANT: FUKUTA, Makoto
;; APPLICANT: IINUMA, Satoshi
;; APPLICANT: OKADA, Hiroaki
;; TITLE OF INVENTION: PEPTIDE CONJUGATE
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: WEGNER, CANTOR, MUELLER & PLAYER
;; STREET: 1233 20th Street, N.W., Suite 300
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20036-8218
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/158,245
;; FILING DATE:
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 318031-1992
;; FILING DATE: 27-NOV-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Cantor, Herbert I.
;; REGISTRATION NUMBER: 24,392
;; REFERENCE/DOCKET NUMBER: P-8700-24068
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 887-0400
;; TELEFAX: (202) 835-0605
;; TELEX: 440706 and 248394
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 21 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-158-245-1

Query Match 50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087; 6; Indels 0; Gaps 0;
Matches 11; Conservative 3; Mismatches 6;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||: 1: 1
DB 2 IVEQCTSCISLYQLENYCN 21

RESULT 6
US-08-233-617-1
;; Sequence 1, Application US/08233617
;; Patent No. 5466666
;; GENERAL INFORMATION:
;; APPLICANT: Obermeier, Rainer
;; APPLICANT: Sabel, Walter
;; APPLICANT: Dell, Peter
;; APPLICANT: Geisen, Karl
;; TITLE OF INVENTION: Amorphous Monospherical Forms of Insulin
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;; ADDRESSEE: Dunner
;; STREET: 1300 I Street, N.W., Suite 700
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005-3315
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/233,617
;; FILING DATE: 25-APR-1994
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: P 43 13 702.4
;; FILING DATE: 27-APR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Carol P. Einaudi
;; REGISTRATION NUMBER: 32,220
;; REFERENCE/DOCKET NUMBER: 02481.1374-00000
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-408-4000
;; TELEFAX: 202-408-4400
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 21 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; ORIGINAL SOURCE:
;; ORGANISM: Escherichia coli
US-08-233-617-1

Query Match 50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087; 6; Indels 0; Gaps 0;
Matches 11; Conservative 3; Mismatches 6;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||: 1: 1
DB 2 IVEQCTSCISLYQLENYCN 21

RESULT 7

US-08-160-376A-1
; Sequence 1, Application US/08160376A
; Patent No. 5473049
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Ranier
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgen
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process For Obtaining Proinsulin
; TITLE OF INVENTION: Possessing Correctly Linked
; TITLE OF INVENTION: Cystine Bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth A. Genoni, Esq.
; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
; CITY: Somerville
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08876-1258
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM 386
; OPERATING SYSTEM: WINDOWS 3.1
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/160,376A
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GE P 4240420.7
; FILING DATE: December 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara V. Maurer, Esq.
; REGISTRATION NUMBER: 31,287
; REFERENCE/DOCKET NUMBER: HOE 92/F 384
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 231-4079
; TELEFAX: (908) 231-2255
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 Amino Acids
; TYPE: Amino Acid
; TOPOLOGY: not relevant
US-08-160-376A-1

Query Match 50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
; IIII II IIII: I: I
Db 2 IVEQCCTSCISLYQLENYCN 21

RESULT 8

US-08-304-070-1
; Sequence 1, Application US/08304070
; Patent No. 5547929
; GENERAL INFORMATION:
; APPLICANT: Anderson Jr., James H.
; APPLICANT: De Felippis, Michael R.
; APPLICANT: Frank, Bruce H.
; APPLICANT: Havel, Henry A.
; TITLE OF INVENTION: INSULIN ANALOG FORMULATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company/Patent Division
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States

ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/304,070
; FILING DATE: 12-SEP-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Caltrider, Steven P.
; REGISTRATION NUMBER: 36,467
; REFERENCE/DOCKET NUMBER: X-9635
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-0757
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-304-070-1

Query Match 50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 2 LVEQASTSQASLYQIYNFDN 21
; IIII II IIII: I: I
Db 2 IVEQCCTSCISLYQLENYCN 21

RESULT 9

US-08-285-661-1
; Sequence 1, Application US/08285661
; Patent No. 5559094
; GENERAL INFORMATION:
; APPLICANT: Brens et. al.
; TITLE OF INVENTION: Asp BInsulin Analogs
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; ADDRESSEE: Patent Division/DKN
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/285,661
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Douglas K. No. 5559094man
; REGISTRATION NUMBER: 33267
; REFERENCE/DOCKET NUMBER: X-9142
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-2958
; TELEFAX: (317) 276-1294
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids


```

; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-285-661-1

Query Match      50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY  2 LVEQASTSQASLYQIYNFDN 21
Db   2 IVEQCCTCSICSLYOLENYCN 21

RESULT 10
US-08-301-838-1
; Sequence 1, Application US/08301838
; Patent No. 5597796
; GENERAL INFORMATION:
; APPLICANT: Brange, Jens J.V.
; TITLE OF INVENTION: TRANSDERMAL INSULIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 55977960 No. 5597796disk of No. 5597796th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/301,838
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/085,473
; FILING DATE: 30-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/751,836
; FILING DATE: 30-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0101/91
; FILING DATE: 22-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK92/00019
; FILING DATE: 22-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3447.210-US
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 6..11
; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 7
; OTHER INFORMATION: /note- "This disulfide bond is
; OTHER INFORMATION: between residue number 7 of this sequence and
; OTHER INFORMATION: residue number 7 of SEQ ID NO:2"
```

```

; FEATURE:
; NAME/KEY: Disulfide-bond
; LOCATION: 20
; OTHER INFORMATION: /note- "This disulfide bond is
; OTHER INFORMATION: between residue number 20 of this sequence and
; OTHER INFORMATION: residue number 19 of SEQ ID NO:2"
US-08-301-838-1

Query Match      50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY  2 LVEQASTSQASLYQIYNFDN 21
Db   2 IVEQCCTCSICSLYOLENYCN 21

RESULT 11
US-08-389-487-4
; Sequence 4, Application US/08389487
; Patent No. 5663291
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgen
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process for Obtaining Insulin Having
; TITLE OF INVENTION: Correctly Linked Cystine Bridges
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,487
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Einandi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481.1424-00000
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-389-487-4

Query Match      50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY  2 LVEQASTSQASLYQIYNFDN 21
Db   2 IVEQCCTCSICSLYOLENYCN 21

RESULT 12
US-08-507-124-2
```

Sequence 2, Application US/08507124
Patent No. 5670618
GENERAL INFORMATION:
APPLICANT: McKenzie, Maureen A.
TITLE OF INVENTION: Insulin-Like Peptide
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/507,124
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/956,342
FILING DATE: 05-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1828-102P
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /label= peptide
OTHER INFORMATION: /note= "Amino terminal sequence of human insulin"
OTHER INFORMATION: A-chain"
US-08-507-124-2

Query Match 50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||:|||||:|:|
Db 2 IVEQCCTSCSLYQLENYCN 21

RESULT 13
US-08-342-931-1
Sequence 1, Application US/08342931
Patent No. 5693609
GENERAL INFORMATION:
APPLICANT: Baker et. al.
TITLE OF INVENTION: Acylated Insulin Analogs
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
ADDRESSEE: Patent Division/SPC
STREET: Lilly Corporate Center

CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/342,931
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Steven P. Caltrider
REGISTRATION NUMBER: 36467
REFERENCE/DOCKET NUMBER: X9720
TELEPHONE: (317) 276-0757
TELEFAX: (317) 277-1917
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: polypeptide
FEATURE:
NAME/KEY: Variable Site
LOCATION: 1
IDENTIFICATION METHOD:
OTHER INFORMATION: "Xaa at position 1 of SEQ ID NO:1 is Gly; or acylated Gly"
US-08-342-931-1

Query Match 50.5%; Score 53; DB 1; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||:|||||:|:|
Db 2 IVEQCCTSCSLYQLENYCN 21

RESULT 14
US-08-508-664-9
Sequence 9, Application US/08508664
Patent No. 5840542
GENERAL INFORMATION:
APPLICANT: Kang, Yup
APPLICANT: YOON, Ji-Won
TITLE OF INVENTION: METHOD FOR MANUFACTURE OF PROINSULIN
TITLE OF INVENTION: WITH HIGH EXPORT YIELD
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: US
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,664
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:

NAME: Ludwig, S. Peter
REGISTRATION NUMBER: 25,351
REFERENCE/DOCKET NUMBER: 0136/08300
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: PROINSULIN A CHAIN
US-08-508-664-9

Query Match 50.5%; Score 53; DB 2; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
Db 2 IVEQCCTSCISLYQLENYCN 21

RESULT 15
US-08-353-476-85
Sequence 85, Application US/08353476
Patent No. 5871902
GENERAL INFORMATION:
APPLICANT: Weininger, Susan
APPLICANT: Weininger, Arthur M
TITLE OF INVENTION: METHOD OF DETECTION OF DNA WITH A
TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st St., Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/353,476
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bencen, Gerard H
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: GP-100
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-353-476-85

Query Match 50.5%; Score 53; DB 2; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
Db 2 IVEQCCTSCISLYQLENYCN 21

RESULT 16
US-08-484-219-1
Sequence 1, Application US/08484219.
Patent No. 5911997
GENERAL INFORMATION:
APPLICANT: Schwabe, Christian
APPLICANT: Unemori, Elaine
TITLE OF INVENTION: RELAXIN-LIKE FACTOR AND METHODS AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,219
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 7842-040
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-484-219-1

Query Match 50.5%; Score 53; DB 2; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
Db 2 IVEQCCTSCISLYQLENYCN 21

RESULT 17
US-08-979-587-1
Sequence 1, Application US/08979587
Patent No. 5922675
GENERAL INFORMATION:
APPLICANT: Baker et. al.
TITLE OF INVENTION: Acylated Insulin Analogs
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
ADDRESSEE: Patent Division/SPC
STREET: Lilly Corporate Center

; NAME/KEY: Protein
; LOCATION: 1..21
US-09-134-836-1

Query Match 50.5%; Score 53; DB 2; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||:1:1
Db 2 IVEQCTSIQSLEYQNYCN 21

RESULT 20

US-08-967-867-2

; Sequence 2, Application US/08967867

; Patent No. 6001604

; GENERAL INFORMATION:

; APPLICANT: HARTMAN, JACOB R.

; APPLICANT: MENDELOVITZ, SIMONA

; APPLICANT: GORECKI, MARIAN

; TITLE OF INVENTION: GENERATION OF HUMAN INSULIN

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: COOPER & DUNHAM LLP

; STREET: 1185 AVENUE OF THE AMERICAS

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/967,867

; FILING DATE: 12-NOV-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/367,454

; FILING DATE: 29-DEC-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: WHITE, JOHN P.

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 41425-A/JPW/GJG

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-278-0400

; TELEFAX: 212-391-0525

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 21 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

US-08-967-867-2

Query Match 50.5%; Score 53; DB 3; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||:1:1
Db 2 IVEQCTSIQSLEYQNYCN 21

RESULT 21

US-08-750-391-3

; Sequence 3, Application US/08750391

; Patent No. 6015713

; GENERAL INFORMATION:

; APPLICANT: Wright Jr., James R.

; APPLICANT: Pohajdak, Bill

; TITLE OF INVENTION: Transgenic Fish in the Treatment of

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Gray Cary Ware & Freidenrich

; STREET: 4365 Executive Drive, Suite 1600

; CITY: San Diego

; STATE: CA

; COUNTRY: USA

; ZIP: 92121

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/750,391

; FILING DATE: 22-MAR-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/CA96/00171

; FILING DATE: 22-MAR-1996

; NAME: Reiter, Stephen E.

; REGISTRATION/DOCKET NUMBER: 31,192

; REFERENCE/DOCKET NUMBER: P41 90379

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 619-677-1409

; TELEFAX: 619-677-1465

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 21 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; ORIGINAL SOURCE:

; ORGANISM: HUMAN INSULIN A CHAIN

US-08-750-391-3

Query Match 50.5%; Score 53; DB 3; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||:1:1
Db 2 IVEQCTSIQSLEYQNYCN 21

RESULT 22

US-09-099-307-1

; Sequence 1, Application US/09099307A

; Patent No. 6221633

; GENERAL INFORMATION:

; APPLICANT: ERTL, JOHANN

; APPLICANT: HABERMANN, PAUL

; APPLICANT: GEISEN, KARL

; APPLICANT: SEIPKE, GERHARD

; TITLE OF INVENTION: NOVEL INSULIN DERIVATIVES HAVING A RAPID ONSET OF ACTION

; FILE REFERENCE: 02481.1597-00000

; CURRENT APPLICATION NUMBER: US/09/099,307A

; CURRENT FILING DATE: 1998-06-18

; EARLIER APPLICATION NUMBER: 19726167.1

; EARLIER FILING DATE: 1997-06-20

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 21

; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-099-307-1

Query Match 50.5%; Score 53; DB 4; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||||:|
Db 2 IVEQCSTCSLYQLENYCN 21

RESULT 23

US-08-900-574-1
; Sequence 1, Application US/089000574
; Patent No. 6221837

; GENERAL INFORMATION:

; APPLICANT: Ertl, Johann

; APPLICANT: Habermann, Paul

; APPLICANT: Geisen, Karl

; APPLICANT: Seipke, Gerhard

; TITLE OF INVENTION: Insulin derivatives with increased zinc

; TITLE OF INVENTION: binding

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett,

; ADDRESSEE: & Dunner, L.L.P.

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: District of Columbia

; COUNTRY: U.S.A.

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/900,574

; FILING DATE: July 24, 1997

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: German Application No. 6221837 19630242.0

; FILING DATE: July 26, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Carol P. Einaudi

; REGISTRATION NUMBER: 32,220

; REFERENCE/DOCKET NUMBER: 02481.1499-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 408-4000

; TELEFAX: (202) 408-4400

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 21 amino acids

; TYPE: Amino acid

; STRANDEDNESS: Single

; TOPOLOGY: linear

; MOLECULE TYPE: Protein

; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli

; FEATURE:

; NAME/KEY: Protein

; LOCATION: 1..21

US-08-900-574-1

Query Match 50.5%; Score 53; DB 4; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||||:|
Db 2 IVEQCSTCSLYQLENYCN 21

RESULT 24

US-08-552-749-1

; Sequence 1, Application US/08552749

; Patent No. 6277816

; GENERAL INFORMATION:

; APPLICANT: Telena, Iilandia

; TITLE OF INVENTION: Human Insulin Analogues

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Zarley, McKee, Thome, Voorhees, & Sease

; STREET: 801 Grand Avenue Suite 3200

; CITY: Des Moines

; STATE: Iowa

; COUNTRY: United States

; ZIP: 50309

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/552,749

; FILING DATE:

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Nebel, Heidi S.

; REGISTRATION NUMBER: 37,719

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 515-288-3667

; TELEFAX: 515-288-1338

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 21 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-552-749-1

Query Match 50.5%; Score 53; DB 4; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 LVEQASTSQASLYQIYNFDN 21
:|||||||:|
Db 2 IVEQCSTCSLYQLENYCN 21

RESULT 25

US-09-386-303A-1

; Sequence 1, Application US/09386303A

; Patent No. 6380355

; GENERAL INFORMATION:

; APPLICANT: Rubroder, Franz-Josef

; Koller, Reinhold

; TITLE OF INVENTION: Improved process for obtaining

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

Mon Dec 23 08:50:59 2002

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/386,303A
; FILING DATE: 31-Aug-1999
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/134,836
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDonnell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..21
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-386-303A-1

Query Match          50.5%; Score 53; DB 4; Length 21;
Best Local Similarity 55.0%; Pred. No. 0.0087;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY      2 LVEQASTSQASLYQIYNFDN 21
       :||| || ||||: | :|
Db       2 IVEQCTSIQSYQLENTCN 21

Search completed: December 23, 2002, 07:17:52
Job time : 11.8824 secs
```